

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 19:08:57 ; Search time 61 Seconds
(without alignments)
666.438 Million cell updates/sec

Title: US-09-308-397-2

Perfect score: 1518

Sequence: 1 MTKTAFLEAGGAGYLGWGR.....QTAHLAHEVDQASIVALLEK 306

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1518	100.0	306	19	FabD polypeptide s
2	1518	100.0	306	19	CPE 34 protein seq
3	1515	99.8	306	19	S. pneumoniae prot
4	1514	99.7	306	22	Streptococcus pneu
5	1048	69.0	308	23	Lactococcus lactis
6	1036	68.2	314	23	Streptococcus poly
7	996.5	65.6	308	23	Streptococcus poly
8	937	61.7	313	22	Enterococcus faec
9	863	56.9	182	19	Carboxy terminal r
10	670	44.1	137	18	S. pneumoniae mal

11	643.5	42.4	309	22	AAU34532	E. coli cellular p
12	636	41.9	312	22	AAU3382	Haemophilus influe
13	613	40.4	122	19	AAW60857	Amino terminal reg
14	588.5	38.8	313	23	ABB48893	Listeria monocytog
15	585	38.5	130	19	AAW61018	Streptococcus pneu
16	556.5	36.7	302	22	AAU33964	Staphylococcus aur
17	556.5	36.7	311	22	AAU36537	Staphylococcus aur
18	556.5	36.7	312	22	AAU36301	Pseudomonas aerugi
19	550.5	36.3	308	21	AAU70134	Staphylococcus aur
20	547.5	36.1	308	19	AAW60032	Amino acid sequenc
21	547.5	36.1	312	22	AAE02192	Staphylococcus aur
22	547.5	36.1	332	22	AAE02191	Staphylococcus aur
23	528	34.8	308	20	AAU34892	Chlamydia pneumoni
24	515.5	34.0	311	23	ABP39979	Staphylococcus epi
25	454.5	29.9	309	20	AAU36958	Protein involved i
26	439.5	29.0	327	21	AAU18524	Arabidopsis thalia
27	439.5	29.0	327	21	AAU18524	Arabidopsis thalia
28	439.5	29.0	367	21	AAU18523	Arabidopsis thalia
29	439.5	29.0	367	21	AAU18523	Arabidopsis thalia
30	439.5	29.0	393	21	AAU18522	Arabidopsis thalia
31	439.5	29.0	393	21	AAU18522	Arabidopsis thalia
32	433	28.5	330	21	AAU07678	Amino acid sequenc
33	419	27.6	3413	19	AAU52849	A. mediterranei xi
34	416.5	27.4	309	22	AAU35697	Helicobacter pylor
35	414.5	27.3	1402	22	AAU81115	Mycobacterium tube
36	414.5	27.3	1402	22	AAU81115	Protein encoded by
37	413.5	27.2	6797	22	AAU31558	Pimaricin biosynth
38	405	26.7	223	22	AAU82280	S. epidermidis ope
39	403	26.5	5435	22	AAE10145	Streptomyces nous
40	399.5	26.3	309	21	AAU47615	Streptomyces nous
41	399.5	26.3	349	21	AAU47615	Arabidopsis thalia
42	399.5	26.3	375	21	AAU47613	Arabidopsis thalia
43	399	26.3	405	21	AAU26155	B. cereus zwittrm
44	397.5	26.2	11096	22	AAE10129	Streptomyces nous
45	379.5	25.0	412	23	AAU77694	Actinomycete monen

ALIGNMENTS

RESULT 1
AAW60856
ID AAW60856 standard; Protein; 306 AA.
AC AAW60856;
DT 24-SEP-1998 (first entry)
DE FabD polypeptide sequence.
KW FabD protein; malonyl-CoA:ACP family; diagnosis; infection; vaccine; screen.
OS Streptococcus pneumoniae.
PN WO9822133-A1.
PD 28-MAY-1998.
PF 14-NOV-1997; 97WO-US20992.
PR 18-NOV-1996; 96US-0031160.
PA (SMK) SMITHKLINE BEECHAM CORP.
PI Gentry DR, Lonsdale JT, Payne DU, Pearson SC, Van Aller G;
DR WPI; 1998-312173/27.
DR N-PSDB; AAU37239.
PT New isolated Streptococcus pneumoniae FabD gene - used to develop products for the diagnosis, prevention and treatment of bacterial diseases, particularly S. pneumoniae infection

XX PS Claim 2; Page 6; 45pp; English.

XX CC The present sequence represents a FabD protein of Streptococcus

CC CC pneumoniae 0105993 (NCIIB 40800). The novel FabD polypeptides are

CC CC related to other proteins of the malonyl-CoA:ACP family. The products

CC CC can be used for the diagnosis of Streptococcus pneumoniae infections.

CC CC Vectors containing the FabD DNA sequence can be administered directly

CC CC to a mammal to produce the FabD peptide to provoke an antibody/T-cell

CC CC response in order to prevent a disease. The peptide can be used to

CC CC screen for compounds which modulate its activity.

XX SQ Sequence 306 AA;

Query Match 100.0%; Score 1518; DB 19; Length 306;

Best Local Similarity 100.0%; Pred. No. 3.2e-131;

Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKTAPFAGGQAQYLGMRDFYDQYPIVKETIDRASQVGLDRLYLIDTEEDKLNQTRY 60

DB 1 MTKTAPFAGGQAQYLGMRDFYDQYPIVKETIDRASQVGLDRLYLIDTEEDKLNQTRY 60

QY 61 TOPAILATSVAIYRLLEQKGYQPDWVAGLSLGEYSALVAGSALDGEDAVALVAKRGAYME 120

DB 61 TOPAILATSVAIYRLLEQKGYQPDWVAGLSLGEYSALVAGSALDGEDAVALVAKRGAYME 120

QY 121 EAPADSGKMWAVLNTPEVIEEACOKASELGWVTPANTPAQIVIAGEVAVDRAVEL 180

DB 121 EAPADSGKMWAVLNTPEVIEEACOKASELGWVTPANTPAQIVIAGEVAVDRAVEL 180

QY 181 LOEAGAKRLIPLKVSQGFHTALLEPASOKLAETLAQVSPSDFTCPLVGNTEAAVMQKEDI 240

DB 181 LOEAGAKRLIPLKVSQGFHTALLEPASOKLAETLAQVSPSDFTCPLVGNTEAAVMQKEDI 240

QY 241 AQLLTQVKPEVPFYESIGVMQEAGISNFIEIGPKVLSGFVKKIDQTAHLAHEVDQASL 300

DB 241 AQLLTQVKPEVPFYESIGVMQEAGISNFIEIGPKVLSGFVKKIDQTAHLAHEVDQASL 300

QY 301 VALLEK 306

DB 301 VALLEK 306

RESULT 2

AAW01031

ID AAW01031 standard; Protein; 306 AA.

XX AC AAW01031;

XX DT 02-OCT-2001 (first entry)

XX CFE 34 protein sequence.

XX Antibacterial; vaccine; gene therapy; bacterial cell wall viability;

XX CFE; CEG; Conserved Essential Gene; bacterial infection;

XX antisense therapy; antibiotic resistance.

XX Streptococcus pneumoniae.

XX WO200149721-A2.

XX 12-JUL-2001.

XX 29-DEC-2000; 2000WO-US35604.

XX 30-DEC-1999; 99US-0174089.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Dougherty TJ, Pucci MJ, Dougherty BA, Davison DB, Bruccoleri RE;

XX Thanassi JA;

XX WPI; 2001-496721/54.

DR N-PSDB; AAW90730.

XX Nucleic acids encoding conserved essential genes involved in bacterial

PT replication which are potential targets for the treatment of antibiotic

PT resistant bacterial infections -

XX Claim 27; Pages 272-273; 380pp; English.

XX The present invention relates to nucleic acids (AAW90701-AAW90918)

CC encoding polypeptides (AAW01002-AAW0114), which are essential for the

CC viability of a bacterial cell wall. The acronym CFE stands for "CEG For

CC Expression", where CEG stands for "Conserved Essential Gene". The nucleic

CC acids are useful for detecting the presence of proteins essential for the

CC viability of a bacterial cell wall in samples such as cells, tissues,

CC biological fluids, blood, serum, nose, ear or throat swabs with ligands,

CC and for detecting corresponding target nucleic acid molecules with

CC complementary sequences. The nucleic acids are also useful for

CC determining whether a genomic nucleotide sequence of interest is

CC essential for viability of a bacterial cell or whether it resides within

CC an operon, by integrating an exogenous nucleotide sequence comprising a

CC portion of an open reading frame of the genomic sequence of interest

CC (comprising 200-500 base pairs) into the genomic sequence of interest

CC which confers a selectable phenotype to the cell, and determining cell

CC viability with a selection agent such as chloramphenicol. The nucleic

CC acids and proteins are also useful as vaccines and for treating bacterial

CC infections with gene therapy and antisense therapy. The nucleic acids

CC also enable identification of targets suitable for the treatment of

CC antibiotic resistant bacterial infections.

XX SQ Sequence 306 AA;

Query Match 100.0%; Score 1518; DB 22; Length 306;

Best Local Similarity 100.0%; Pred. No. 3.2e-131;

Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKTAPFAGGQAQYLGMRDFYDQYPIVKETIDRASQVGLDRLYLIDTEEDKLNQTRY 60

DB 1 MTKTAPFAGGQAQYLGMRDFYDQYPIVKETIDRASQVGLDRLYLIDTEEDKLNQTRY 60

QY 61 TOPAILATSVAIYRLLEQKGYQPDWVAGLSLGEYSALVAGSALDGEDAVALVAKRGAYME 120

DB 61 TOPAILATSVAIYRLLEQKGYQPDWVAGLSLGEYSALVAGSALDGEDAVALVAKRGAYME 120

QY 121 EAPADSGKMWAVLNTPEVIEEACOKASELGWVTPANTPAQIVIAGEVAVDRAVEL 180

DB 121 EAPADSGKMWAVLNTPEVIEEACOKASELGWVTPANTPAQIVIAGEVAVDRAVEL 180

QY 181 LOEAGAKRLIPLKVSQGFHTALLEPASOKLAETLAQVSPSDFTCPLVGNTEAAVMQKEDI 240

DB 181 LOEAGAKRLIPLKVSQGFHTALLEPASOKLAETLAQVSPSDFTCPLVGNTEAAVMQKEDI 240

QY 241 AQLLTQVKPEVPFYESIGVMQEAGISNFIEIGPKVLSGFVKKIDQTAHLAHEVDQASL 300

DB 241 AQLLTQVKPEVPFYESIGVMQEAGISNFIEIGPKVLSGFVKKIDQTAHLAHEVDQASL 300

QY 301 VALLEK 306

DB 301 VALLEK 306

RESULT 3

AAW80671

ID AAW80671 standard; Protein; 306 AA.

XX AC AAW80671;

XX DT 24-DEC-1998 (first entry)

XX S. pneumoniae protein (ACPtrans).

XX Streptococcus pneumoniae protein; recombinant; gene expression; DNA chip;

XX virulence; antibody; infection; detection; treatment; hypothetical;

XX cell wall biosynthetic; external target; minimal gene set protein.

XX Streptococcus pneumoniae.
 OS WO9826072-A1.
 XX 18-JUN-1998.
 XX 09-DEC-1997; 97WO-US22578.
 XX 13-DEC-1996; 96US-0036281.
 XX (ELIL) LILLY & CO ELI.
 XX Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR;
 PI Mills BJ, Norris FH, Peery RB, Rockey PK, Rostek PR;
 PI Skatrud PL, Smith MC, Solenberg PU, Treadway PU;
 PI Young Bellido ML;
 XX WPI; 1998-348529/30.
 XX Streptococcus pneumoniae nucleic acid sequences - used in DNA chips
 PT for evaluating gene expression, and identification of virulence
 PT genes
 XX Claim 3; Pages 259-259; 333pp; English.
 XX This sequence represents a Streptococcus pneumoniae protein. The
 CC invention provides DNA sequences (AAV65201 to AAV65304) from the
 CC Streptococcus pneumoniae genome and corresponding protein sequences
 CC (AAW80605 to AAW80728). The protein sequences are classified as
 CC hypothetical, cell wall biosynthetic, external target, or minimal gene
 CC set proteins. A recombinant host containing a vector comprising any of
 CC the above nucleic acids can be used for the recombinant expression of the
 CC proteins. The invention also provides a DNA chip having arrayed on it at
 CC least 15 base pair fragment of any one or more of these DNA sequences.
 CC The DNA chip can be used methods for evaluating gene expression in S.
 CC pneumoniae and for identifying virulence genes in S. pneumoniae.
 CC Antibodies that selectively bind to the above proteins or peptide
 CC fragments can be used to treat S. pneumoniae infection. The antibodies
 CC can also be used to detect S. pneumoniae cells.
 XX SQ Sequence 305 AA;
 Query Match 99.8%; Score 1515; DB 19; Length 306;
 Best Local Similarity 99.7%; Pred. No. 6.1e-131;
 Matches 305; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTKTFLFAGGGAQYLGMRDFDYQYPIVKETIDRASQVGLGYDLRYLIDTEEDKLNTRY 60
 DB 1 MTKTFLFAGGGAQYLGMRDFDYQYPIVKETIDRASQVGLGYDLRYLIDTEEDKLNTRY 60
 QY 61 TOPAILATSVATYRLLOEKGYQPMVAGLSLGEYSALVAGSALFDPAVALVAKRGAYME 120
 DB 61 TOPAILATSVATYRLLOEKGYQPMVAGLSLGEYSALVAGSALFDPAVALVAKRGAYME 120
 QY 121 EAPADSGKMWAVLTPVEIEEACOKASELGVVTPANNYTPAQIVLAGEVAVDRAVEL 180
 DB 121 EAPADSGKMWAVLTPVEIEEACOKASELGVVTPANNYTPAQIVLAGEVAVDRAVEL 180
 QY 181 LOEAGAKLIPKVSQPHHTALLEPASQKLAETLAQVSFSDFTPLVGNTEAAVWQKEDI 240
 DB 181 LOEAGAKLIPKVSQPHHTALLEPASQKLAETLAQVSFSDFTPLVGNTEAAVWQKEDI 240
 QY 241 AQLLTRQVKEPVRVYESIGVMOEAGISNFIIEGPKVLSGFVKIKIDQTAHLAHVEDQASL 300
 DB 241 AQLLTRQVKEPVRVYESIGVMOEAGISNFIIEGPKVLSGFVKIKIDQTAHLAHVEDQASL 300
 QY 301 VALLEK 306
 DB 301 VALLEK 306

RESULT 4

AAU37987
 ID AAU37987 standard; Protein; 306 AA.
 XX AC AAU37987;
 XX 14-FEB-2002 (first entry)
 XX Streptococcus pneumoniae cellular proliferation protein #416.
 DE Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 XX Streptococcus pneumoniae.
 OS WO200170955-A2.
 XX 27-SEP-2001.
 XX 21-MAR-2001; 2001WO-US09180.
 XX 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX (ELIT-) ELITRA PHARM INC.
 PA Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX WPI; 2001-611495/70.
 DR N-PSDB; AAG55846.
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX Example 3; Seq ID No 13580; 511pp; English.
 XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes themselves and the discovery of novel antibiotics, the essential
 CC genes, themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 306 AA;
 Query Match 99.7%; Score 1514; DB 22; Length 306;
 Best Local Similarity 99.7%; Pred. No. 7.5e-131;
 Matches 305; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTKTFLFAGGGAQYLGMRDFDYQYPIVKETIDRASQVGLGYDLRYLIDTEEDKLNTRY 60
 DB 1 MTKTFLFAGGGAQYLGMRDFDYQYPIVKETIDRASQVGLGYDLRYLIDTEEDKLNTRY 60
 QY 61 TOPAILATSVATYRLLOEKGYQPMVAGLSLGEYSALVAGSALFDPAVALVAKRGAYME 120

Db 61 TOPAILATSVAIYRLLOEKGYOPDMVAGLSGEYSALVAGSALDFEDAVAVAKRGAYME 120
 QY 121 EAAPADSGKWAVLNTPEVEIEACQKASELGVVTPANTPAQIVIAGEVAVDRAVEL 180
 Db 121 EAAPADSGKWAVLNTPEVEIEACQKASELGVVTPANTPAQIVIAGEVAVDRAVEL 180
 QY 181 LOEAGAKRLIPLKVGSPHTALLPASPQKLAETLAQVFSDFCTPLVGNTEAAVMQKEDI 240
 Db 181 LOEAGAKRLIPLKVGSPHTALLPASPQKLAETLAQVFSDFCTPLVGNTEAAVMQKEDI 240
 QY 241 AQLLTROVKEPVRVYESIGVMQEAGISNFIETIGPGKVLGSGFVKKIDOTAHLAHVDDQASL 300
 Db 241 AQLLTROVKEPVRVYESIGVMQEAGISNFIETIGPGKVLGSGFVKKIDOTAHLAHVDDQASL 300
 QY 301 VALLEK 306
 Db 301 VALLEK 306

RESULT 5
 ABB54086
 ID ABB54086 standard; Protein; 308 AA.
 AC ABB54086;
 XX
 DT 16-MAY-2002 (first entry)
 XX
 DE Lactococcus lactis protein fabd.
 XX
 KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
 XX
 OS Lactococcus lactis ILL1403.
 XX
 PN FR2807446-A1.
 XX
 PD 12-OCT-2001.
 XX
 PF 11-APR-2000; 2000FR-0004630.
 XX
 PR 11-APR-2000; 2000FR-0004630.
 XX
 PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
 XX
 PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
 XX
 DR WPI; 2002-043418/06.
 XX
 PT New nucleotide sequence useful in the identification of Lactococcus
 PT lactis and related species -
 XX
 PS Claim 6; SEQ ID No 789; 2504pp; French.
 XX
 CC The present invention is related to a Lactococcus lactis nucleotide
 CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
 CC nucleic acid sequence is useful in the detection and/or amplification of
 CC nucleic acid sequence, particularly to identify Lactococcus lactis or
 CC related species. The proteins of the invention are useful for the
 CC biosynthesis or biodegradation of a composition of interest. The
 CC invention helps research in lactic bacteria, particularly useful in the
 CC production of yogurt and cheese.
 CC Note: The sequence data for this patent is based on equivalent patent
 CC WO2001/7334 (published 18-OCT-2001) which is available in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 CC
 SQ Sequence 308 AA;
 Query Match 69.0%; Score 1048; DB 23; Length 308;
 Best Local Similarity 67.5%; Pred. No. 5.4e-88;
 Matches 208; Conservative 40; Mismatches 58; Indels 2; Gaps 1;
 QY 1 MTKTAFLEACQAGYLGMRDFFDQYPTVKEVETIDRASQVLYGDLRLVLIETEDKLNQTRY 60
 Db 1 MTKTAFLEACQAGYLGMRDFFDQYPTVKEVETIDRASQVLYGDLRLVLIETEDKLNQTRY 60

QY 61 TOPAILATSVAIYRLLOEKGYOPDMVAGLSGEYSALVAGSALDFEDAVAVAKRGAYME 120
 Db 61 TOPAILATSVAIYRLLOEKGYOPDMVAGLSGEYSALVAGSALDFEDAVAVAKRGAYME 120
 QY 121 EAAPADSGKWAVLNTPEVEIEACQKASELGVVTPANTPAQIVIAGEVAVDRAVEL 178
 Db 121 EAAPADSGKWAVLNTPEVEIEACQKASELGVVTPANTPAQIVIAGEVAVDRAVEL 180
 QY 179 ELIQAAGAKRLIPLKVGSPHTALLPASPQKLAETLAQVFSDFCTPLVGNTEAAVMQKEDI 238
 Db 181 ELIQAAGAKRLIPLKVGSPHTALLPASPQKLAETLAQVFSDFCTPLVGNTEAAVMQKEDI 240
 QY 239 DIAQLLTROVKEPVRVYESIGVMQEAGISNFIETIGPGKVLGSGFVKKIDOTAHLAHVDDQASL 298
 Db 241 DIAQLLTROVKEPVRVYESIGVMQEAGISNFIETIGPGKVLGSGFVKKIDOTAHLAHVDDQASL 300
 QY 299 SLVALLEK 306
 Db 301 SPEALINQ 308

RESULT 6
 ABP28010
 ID ABP28010 standard; Protein; 314 AA.
 AC ABP28010;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Streptococcus polypeptide SEQ ID NO 5196.
 XX
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW Group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; infection; vaccine; meningitis; Gene therapy.
 XX
 OS Streptococcus pyogenes.
 XX
 PN WO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-GB04789.
 XX
 PR 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
 PI Tettelein H;
 XX
 DR WPI; 2002-352536/38.
 DR N-PSDB; ABB68641.
 XX
 PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 PS Claim 1; Page 3863; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC Streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABB68641-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to

CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX
 SQ Sequence 314 AA;
 Query Match 68.2%; Score 1036; DB 23; Length 314;
 Best Local Similarity 68.3%; Pred. No. 7e-87;
 Matches 209; Conservative 39; Mismatches 58; Indels 0; Gaps 0;
 QY 1 MTKTFLPAGQAQYLGMRDFYDQYPIVKETIDRASQVLGVDLYLIDTEEDKLNQTRY 60
 Db 3 MTKTFLPAGQAQYLGMRDFYDQYPIVKETIDRASQVLGVDLYLIDTEEDKLNQTRY 62
 QY 61 TOPAILATSVAIYRL-LQEKGYOPDMVAGLSLGEYSALVAGSALDFEDAVALVAKRGAYME 120
 Db 63 TOPAILATSVAIYRL-LQEKGYOPDMVAGLSLGEYSALVAGSALDFEDAVALVAKRGAYME 122
 QY 121 EAPADSGKQVAVLNTPEVEIEACOKASELGVVTPANYNTPAQIVTAGEVAVVDRVAVEL 180
 Db 123 EAPADSGKQVAVLNTPEVEIEACOKASELGVVTPANYNTPAQIVTAGEVAVVDRVAVEL 182
 QY 181 LQEKAGKRLIPLKVSQVPGHTALLPASPQKLAETLAQVSFSDFTCPVLGNTAEAAVWQKEDI 240
 Db 183 LKRGVKKLILPLNVSGPHTALLPASPQKLAETLAQVSFSDFTCPVLGNTAEAAVWQKEDI 242
 QY 241 AQLLTROYKEPVRYESIGVWQEGISNFIIEGPKVLSGFVKKIDQTAHLAHVEDQASL 300
 Db 243 PELLARQWNEPVRYESIGVWQEGISNFIIEGPKVLSGFVKKIDQTAHLAHVEDQASL 302
 QY 301 VALLEK 306
 Db 303 RLFLDR 308
 RESULT 7
 ABP28009
 ID ABP28009 standard; Protein; 308 AA.
 XX
 AC ABP28009;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Streptococcus polypeptide SEQ ID NO 5194.
 XX
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antinflammatory; infection; vaccine; meningitis; gene therapy.
 XX
 OS Streptococcus agalactiae.
 XX
 WO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-GB04789.
 XX
 PR 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
 PI Tettelin H;
 XX
 WPI; 2002-352536/38.
 DR N-PSDB; ABN68640.

XX New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 PS Claim 1; Page 3863; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (51), given in
 CC the specification. The proteins have antibacterial and antinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX
 SQ Sequence 308 AA;
 Query Match 65.6%; Score 996.5; DB 23; Length 308;
 Best Local Similarity 66.3%; Pred. No. 2.9e-83;
 Matches 203; Conservative 39; Mismatches 63; Indels 1; Gaps 1;
 QY 1 MTKTFLPAGQAQYLGMRDFYDQYPIVKETIDRASQVLGVDLYLIDTEEDKLNQTRY 60
 Db 1 MNKVSFLPAGQAQYLGMRDFYDQYPIVKETIDRASQVLGVDLYLIDTEEDKLNQTRY 60
 QY 61 TOPAILATSVAIYRL-LQEKGYOPDMVAGLSLGEYSALVAGSALDFEDAVALVAKRGAYM 119
 Db 61 TOPAILATSVAIYRL-LQEKGYOPDMVAGLSLGEYSALVAGSALDFEDAVALVAKRGAYM 120
 QY 120 EAPADSGKQVAVLNTPEVEIEACOKASELGVVTPANYNTPAQIVTAGEVAVVDRVAVEL 179
 Db 121 EAPADSGKQVAVLNTPEVEIEACOKASELGVVTPANYNTPAQIVTAGEVAVVDRVAVEL 180
 QY 180 LQEKAGKRLIPLKVSQVPGHTALLPASPQKLAETLAQVSFSDFTCPVLGNTAEAAVWQKEDI 239
 Db 181 ELKQGVKKLILPLNVSGPHTALLPASPQKLAETLAQVSFSDFTCPVLGNTAEAAVWQKEDI 240
 QY 240 AQLLTROYKEPVRYESIGVWQEGISNFIIEGPKVLSGFVKKIDQTAHLAHVEDQAS 299
 Db 241 IKSLLARQWNEPVRYESIGVWQEGISNFIIEGPKVLSGFVKKIDQTAHLAHVEDQAS 300
 QY 300 LVALLE 305
 Db 301 FNNLKE 306
 RESULT 8
 AAU35327
 ID AAU35327 standard; Protein; 313 AA.
 XX
 AC AAU35327;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE Enterococcus faecalis cellular proliferation protein #614.
 XX
 KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 XX
 OS Enterococcus faecalis.
 XX
 PN WO200170955-A2.
 XX

ID XX AA011297 standard; Protein; 137 AA.
 AC AA011297;
 DT 20-MAY-1999 (first entry)
 XX S. pneumoniae malonyl CoA-acyl carrier protein transacylase.
 DE Streptococcus pneumoniae strain 0100993; vaccine; immune response;
 XX streptococcal infection; pneumococcal.
 OS Streptococcus pneumoniae.
 XX WO9737026-A1.
 XX 09-OCT-1997.
 XX 01-APR-1997; 97WO-US05306.
 XX 22-AUG-1996; 96US-0025788.
 XX 02-APR-1996; 96US-0014690.
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 XX (SMIK) SMITHKLINE BEECHAM PLC.
 PA Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
 PI Stodola RK;
 XX WPI; 1997-503111/46.
 DR N-PSDB; AAX30880.
 XX Nucleic acids encoding pneumococcal polypeptide(s) - useful in
 PT vaccines, drug screening, etc
 XX Claim 6; Page 305; 354pp; English.
 CC AAX30724 to AAX30946 represent genomic DNA sequences isolated from
 CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences
 CC encode the novel proteins given in AAX1114 to AAX11367. The proteins,
 CC isolated from Streptococcus pneumoniae, can be used in vaccines against
 CC streptococcal infections and in assays for identifying compounds that
 CC inhibit or activate the activity of the proteins. The antagonists can
 CC be used to treat an individual having need to inhibit a bacterial
 CC protein. Vectors expressing the proteins can be used to induce a
 CC protective immune response in mammals.
 XX Sequence 137 AA;
 SQ
 Query Match 44.1%; Score 670; DB 18; Length 137;
 Best Local Similarity 97.8%; Pred. No. 9.8e-54;
 Matches 134; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MTKTAFAGGGAQYLGMRDFDYQYPIVKETIDRASQVLGYDLRYLIDTEEDKLNQTRY 60
 DB 1 MTKTAFAGGGAQYLRMGDFDYQYPIVKETIDRASQVLGYDLRYLIDTEEDKLNQTRY 60
 QY 61 TOPAILATSVAIYRLLOEKGYQPDWVAGISLGEYSALVSGALDFEDAVLVAKRGAYNE 120
 DB 61 TOPAILATSVAIYRLLOEKGYHDPWVAGISLGEYSALVSGALDFEDAVLVAKRGAYNE 120
 QY 121 EAAPADSGKXMAVINTP 137
 DB 121 EAAPADSGKXMAVLHTP 137
 RESULT 11
 AAU34532
 ID AAU34532 standard; Protein; 309 AA.
 AC AAU34532;
 XX 14-FEB-2002 (first entry)
 DT
 XX

DE E. coli cellular proliferation protein #113.
 XX Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 XX Escherichia coli.
 OS WO200170955-A2.
 XX 27-SEP-2001.
 PD 21-MAR-2001; 2001WO-US09180.
 XX 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-208848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX (ELIT-) ELITRA PHARM INC.
 PA Haselbeck R, Ohlsen XL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 XX Yamamoto RT, Xu HH;
 PI WPI; 2001-611495/70.
 XX N-PSDB; AAS52391.
 DR New polynucleotides for the identification and development of
 XX antibiotics, comprise sequences of antisense nucleic acids -
 PT Example 3; Seq ID No 10125; 511pp; English.
 XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 309 AA;
 SQ
 Query Match 42.4%; Score 643.5; DB 22; Length 309;
 Best Local Similarity 45.5%; Pred. No. 8.5e-51;
 Matches 141; Conservative 60; Mismatches 102; Indels 7; Gaps 4;
 QY 1 MTKTAFAGGGAQYLGMRDFDYQYPIVKETIDRASQVLGYDLRYLIDTEEDKLNQ 57
 DB 1 MTQFAFVPPGGSGQTGVNLMADMAASYPIVETFAESAALGYDLWAL--TQQGPAEELNK 58
 QY 58 TRYTOPAILATSVAIYRLLOEKGYQ-PDWVAGISLGEYSALVSGALDFEDAVLVAKRG 116
 DB 59 TWQTOPALLTASVALYRWVQQGGKAPAMWAGHSLGEYSALVCGAVIDFADAVLVMRG 118
 QY 117 AYMEERAPADSGKXMAVINTPVEVIEEACQKASELGVVTPANTPAQIIVAGEVAVDR 176
 DB 119 KFMQEAIVPEGTGMAAAIIGLDDASIAKAEAAEQGVVSPVNFNSPGQVVIAGKEAVER 178
 QY 177 AVELLQAGAKRLIPLKVSQGFHTALDEPASQKLAETLAQVSPSDFTCPLVGNTEAAVMQ 236

Db 179 AGAACAAAGAKRALPLFVSFVSHCALMKPAADKLAVALAKITFNAPTVPVNNVDVKET 238
 QY 237 KED-IAQLTRQVKEPVRFYFESIGVMQFAGISNFTIEIGPGKVLGSGFKKKIDQTAHLAHE 295
 Db 239 NGDAIRDALVQLYNFQWTKSVEMAAQGVHELVYEVGPGKVLGLTKRIVDTITASALN 298
 QY 296 DOASLVALLE 305
 Db 299 EFSMAAALE 308

RESULT 12
 AAU35382
 ID AAU35382 standard; Protein; 312 AA.
 XX
 AC AAU35382;
 DT 14-FEB-2002 (first entry)
 XX
 DE Haemophilus influenzae cellular proliferation protein #23.
 XX
 KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 XX
 OS Haemophilus influenzae.
 XX
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US09180.
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 DR WPI; 2001-611495/70.
 DR N-PSDB; AAS3241.
 XX
 XX New polynucleotides for the identification and development of
 PT antibiotic, comprise sequences of antisense nucleic acids -
 XX
 PS Example 3; Seq ID No 10975; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format-directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 312 AA;
 Query Match 41.9%; Score 636; DB 22; Length 312;
 Best Local Similarity 46.0%; Pred. No. 4.2e-50;
 Matches 142; Conservative 54; Mismatches 109; Indels 4; Gaps 3;
 QY 1 MYKTAFLPAGQAQVGLNGRDFYDQPIVKETIDRASQVGLGYDLRYLIDT-BEDKLNQTR 59
 Db 1 MKKFAWVFGQSGTQVGLADLATEPIVIEFKQASDALGYDLWLVQQGFABELNKTW 60
 QY 60 YTPQAILATSVAIYELLOEKGYQ--PDMVAGLSLGEYSALVASGALDFEDAVLVAKRGA 117
 Db 61 QTPALLAASVAIVYRWKEKFPOLKPEVWAGHSLSGEYSALVCAGVLDQDAIKLVLRGK 120
 QY 118 YMEEAAPADSGRMVAVLNTPEVIEBACQKASELGVTPANVTQAQIVIAGEVAVDRA 177
 Db 121 LMQQAVPEGTGAMYAIIGLDNEAIIINACKQABEGEVSAVNFNSPGQVVIAGAKAVERA 180
 QY 178 VELLOFAGAKRLIPLKVGSPFHTALLEPASOKIASTLAQVSFSDTCTPLVGNTEA-AVMQ 236
 Db 181 AALCKEAGAKRALPLAVSVPSHCALMKPAEQLAVTLENIQINTTISVLNNVDVKAETE 240
 QY 237 KEDIAQLTRQVKEPVRFYFESIGVMQFAGISNFTIEIGPGKVLGSGFKKKIDQTAHLAHE 296
 Db 241 GTEIRTAIVRQLYSPVRMTETVERKMQDGLVLAESVGPGLVGLTKRIVGDLQAISSVD 300
 QY 297 QASLVALLE 305
 Db 301 VASFNAVEE 309

RESULT 13
 AAU60857
 ID AAU60857 standard; Protein; 122 AA.
 XX
 AC AAU60857;
 DT 24-SEP-1998 (first entry)
 XX
 DE Amino terminal region of a FabD polypeptide embodiment.
 XX
 KW FabD protein; malonyl-CoA:ACP family; diagnosis; infection; vaccine;
 KW screen.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO9822133-A1.
 XX
 PD 28-MAY-1998.
 XX
 PF 14-NOV-1997; 97WO-US20992.
 XX
 PR 18-NOV-1996; 96US-0031160.
 XX
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 PA
 XX Gentry DR, Lonsdale JT, Payne DJ, Pearson SC, Van Aller G;
 PI
 XX WPI; 1998-312173/27.
 DR N-PSDB; AAV37240.
 XX
 XX New isolated Streptococcus pneumoniae FabD gene - used to develop
 PT products for the diagnosis, prevention and treatment of bacterial
 PT diseases, particularly S. pneumoniae infection
 XX
 PS Claim 23; Page 7; 45pp; English.
 XX
 CC The present sequence represents the amino terminal region of a FabD
 CC embodiment of Streptococcus pneumoniae 0100993 (NCIMB 40800). The novel
 CC FabD polypeptides are related to other proteins of the malonyl-CoA:ACP
 CC family. The products can be used for the diagnosis of Streptococcus
 CC pneumoniae infections. Vectors containing the FabD DNA sequence can be

CC administered directly to a mammal to produce the FabD peptide to
 CC provoke an antibody/T-cell response in order to prevent a disease. The
 CC Peptide can be used to screen for compounds which modulate its activity.
 XX Sequence 122 AA;
 Query Match 40.4%; Score 613; DB 19; Length 122;
 Best Local Similarity 100.0%; Pred. No. 1.5e-48;
 Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTKTAFIFAGGAQVGLGMRDFYDQYPIVKTIDRASQVLYDLYLIDTEEDKLNQTRY 60
 DB 1 MTKTAFIFAGGAQVGLGMRDFYDQYPIVKTIDRASQVLYDLYLIDTEEDKLNQTRY 60
 QY 61 TOPAILATSVIYRLLOEKGYQPDWAGLSLGEYSALVASCALDFEDAVALKRGAYME 120
 DB 61 TOPAILATSVIYRLLOEKGYQPDWAGLSLGEYSALVASCALDFEDAVALKRGAYME 120
 QY 121 EA 122
 DB 121 EA 122
 RESULT 14
 ID ABB48893 standard; Protein; 313 AA.
 AC ABB48893;
 XX
 DT 05-FEB-2002 (first entry)
 XX
 DE Listeria monocytogenes protein #1597.
 XX
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KW vitamin B12; bacterial infection; disease.
 XX
 OS Listeria monocytogenes.
 XX
 PN WO200177335-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 11-APR-2001; 2001WO-FR01118.
 XX
 PR 11-APR-2000; 2000FR-0004629.
 XX
 FA (INSP) INST PASTEUR.
 XX
 PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Feihi H, Dehoux P,
 PI Dusserget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;
 PI Danieles J, Goebel W, Kreft J, Kuhn M, Ng B, Vazquez-Boland JA;
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
 PI Rose M, Voss H;
 XX
 DR WPI; 2002-010914/01.
 XX
 PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 PT and prevention of Listeria and related bacterial infections, and
 PT related polypeptides -
 XX
 PS Claim 6; SEQ ID No 1598; 192pp; French.
 XX
 CC The present invention relates to the genome sequence of Listeria
 CC monocytogenes ED-e (see ABA03041). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in L.
 CC monocytogenes and related organisms, and for studying genetic
 CC polymorphisms and other genomes. The present invention is a protein
 CC encoded by the genome sequence of the present invention. Proteins
 CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of L. monocytogenes and related organisms, and

CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate L. monocytogenes-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccine compositions for the treatment or prevention of infections by L.
 CC monocytogenes and related organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 313 AA;
 Query Match 38.8%; Score 588.5; DB 23; Length 313;
 Best Local Similarity 43.1%; Pred. No. 9.9e-46;
 Matches 131; Conservative 58; Mismatches 110; Indels 5; Gaps 4;
 QY 1 MTKTAFIFAGGAQVGLGMRDFYDQYPIVKTIDRASQVLYDLYLIDTES--DKLNQ 58
 DB 1 MTKIAFPVFGGAQVGLGMRDFYDQYPIVKTIDRASQVLYDLYLIDTES--DKLNQ 58
 QY 59 RYTOPAILATSVIYRLLOEKGYQPDWAGLSLGEYSALVASCALDFEDAVALKRGAY 118
 DB 60 ENAQPALVSTSVAILRALETYGVKADYVAGHSLGEYSALVAGGFLEASDAIYLVKRGEL 119
 QY 119 MEERAPADSGKMAVAVLNTPEVIEEACOKASBLG-VVTPANVNTPAQIVIAEVAVDRA 177
 DB 120 MEAAVNGAGAAVAVLGVDRRLTKITTEVTKEGDAVQANLNCQGIIVISGTTAGVEKA 179
 QY 178 VELLOBAGAKRLIPUKVSGPHTALLEPASQKLAETLACQVFSDFTCPLVGNTEA-AMQ 236
 DB 180 GEKAKESGAKRVLPVAVSGPHTALLEPASQKLAETLACQVFSDFTCPLVGNTEA-AMQ 239
 QY 237 KEDIAQLLTQVKEPVRVFSIGVMQEGAGISNFIIGPKVLSGFVKKIDQTAHLAHVED 296
 DB 240 KSEISDKLIKQIYSPVLWEDIVEELIKNGVDTFVEIGSKVLGLIKKINRDVTVLSAGD 299
 QY 297 QASL 300
 DB 300 AHSV 303
 RESULT 15
 AA61018
 ID AA61018 standard; Protein; 130 AA.
 XX
 AC AA61018;
 XX
 DT 13-OCT-1998 (first entry)
 XX
 DE Streptococcus pneumoniae encoded polypeptide.
 XX
 KW coding region; ORF; open reading frame; antibacterial;
 KW infection; prevention; meningitis.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO9819689-A1.
 XX
 PD 14-MAY-1998.
 XX
 PF 27-OCT-1997; 97WO-US19226.
 XX
 PR 01-NOV-1996; 96US-0029930.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 PI Black MT, Hodgson JS, Knowles DJC, Lonetto MA, Nicholas RO;
 PI Reid RH, Zarfos FN;
 XX
 DR WPI; 1998-286586/25.
 DR N-PSDB; AAV37406.

```

XX      New isolated nucleic acids from Streptococcus pneumoniae - useful,
PT      e.g. for identifying anti-bacterial(s) for treatment and prevention
XX      of meningitis
XX
XX      Claim 11; Page 112; 130pp; English.
XX
XX      The sequence is that of the polypeptide encoded by a region isolated
CC      from S. pneumoniae which shows homology to malonyl coenzyme A-acyl
CC      carrier protein transacylase. It, or agonists of it, may be useful as an
CC      antibacterial for treatment or prevention of infection, specifically
CC      caused by S.pneumoniae (particularly meningitis) but possibly also
CC      Helicobacter pylori (ulcers and gastric cancer). It may be of particular
CC      use before insertion of an in-dwelling device or any other
CC      invasive procedure. The protein, or nucleic acid encoding
CC      it, can also be used in vaccines to induce a cellular
CC      and/or humoral immune response, or to screen for other
CC      antibacterials. The DNA may also contain flanking sequences
CC      that are potential sources of control elements for bacterial
CC      gene expression. Detecting a sequence encoding the protein
CC      can be used diagnostically, e.g. to detect a mutation for
CC      serotyping or classifying infectious agents.
XX
XX      Sequence 130 AA;
SQ
      Query Match      38.5%; Score 585; DB 19; Length 130;
      Best Local Similarity 93.8%; Pred. No. 6e-46;
      Matches 121; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY      153 VTPANYNTPAQIIVAGIEVAVDRAVELLOEAGAKGLIELKVGSPFHTALLEPASOKLAE 212
      |||||
DB      1 VTPANYNTPAQIIVAGIEVAVDRAVELLOEAGAKGLIELKVGSPFHTALLEPASOKLAE 60
      |||||

QY      213 TLAQVFSDFTCPLVGNTEAAVWQKEDIQALLTRQVKPEVRYESIGVWQAGISNFTFI 272
      |||||
DB      61 TLAQVFSDFTCPLVGNTEAAVWQKEDIQALLTRQVKPEVRYESIGVWQAGISNFTFI 120
      |||||

QY      273 GPGKVLSCF 281
      : : |
DB      121 WTGESLVRF 129

```

Search completed: June 11, 2003, 20:02:50
Job time : 63 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 20:01:47 ; Search time 34 Seconds
(without alignments)
264.806 Million cell updates/sec

Title: US-09-308-397-2
Perfect score: 1518
Sequence: 1 MTKTAFIFAGQGAQYLGMGR.....QTAHLAHVEDQASVALLK 306

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pap.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pap.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pap.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pap.*
5: /cgn2_6/ptodata/2/iaa/PTUS COMB.pap.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	547.5	36.1	308	2	US-08-789-609A-2
2	547.5	36.1	308	4	US-09-108-517-2
3	515.5	34.0	311	4	US-09-134-001C-4824
4	376.5	24.8	4150	4	US-09-428-517-2
5	374	24.6	2259	4	US-09-413-814-70
6	374	24.6	2439	3	US-09-335-409-7
7	374	24.6	2439	4	US-09-568-102-7
8	374	24.6	2439	4	US-09-567-969-7
9	374	24.6	2439	4	US-09-568-480-7
10	374	24.6	2439	4	US-09-568-486-7
11	374	24.6	2439	4	US-09-568-472-7
12	374	24.6	2439	4	US-09-567-899-7
13	374	24.6	2475	4	US-09-413-814-48
14	372.5	24.5	5215	4	US-09-105-537-2
15	367.5	24.2	1864	2	US-08-804-227C-3
16	356	23.5	7257	3	US-09-335-409-5
17	356	23.5	7257	4	US-09-568-102-5
18	356	23.5	7257	4	US-09-567-969-5
19	356	23.5	7257	4	US-09-568-480-5
20	356	23.5	7257	4	US-09-568-486-5
21	356	23.5	7257	4	US-09-568-472-5
22	356	23.5	7257	4	US-09-567-899-5
23	351	23.1	1421	3	US-09-335-409-2
24	351	23.1	1421	4	US-09-568-102-2
25	351	23.1	1421	4	US-09-567-969-2
26	351	23.1	1421	4	US-09-568-480-2
27	351	23.1	1421	4	US-09-568-486-2

28	351	23.1	1421	4	US-09-568-472-2	Sequence 2, Appli
29	351	23.1	1421	4	US-09-567-899-2	Sequence 2, Appli
30	350.5	23.1	1841	2	US-08-804-227C-6	Sequence 6, Appli
31	349.5	23.0	6095	4	US-09-144-085-2	Sequence 2, Appli
32	348.5	23.0	2756	1	US-08-375-709-11	Sequence 11, Appli
33	348.5	23.0	2756	1	US-08-752-929-11	Sequence 11, Appli
34	348.5	23.0	2756	4	US-09-090-793-7	Sequence 7, Appli
35	344.5	22.7	4928	4	US-09-036-987A-5	Sequence 5, Appli
36	344.5	22.7	4928	4	US-09-370-700-5	Sequence 5, Appli
37	339.5	22.4	5588	4	US-09-036-987A-6	Sequence 6, Appli
38	339.5	22.4	5588	4	US-09-370-700-6	Sequence 6, Appli
39	338.5	22.3	5087	4	US-09-144-085-1	Sequence 1, Appli
40	334	22.0	343	3	US-08-858-003-32	Sequence 32, Appli
41	334	22.0	343	3	US-09-078-166-32	Sequence 32, Appli
42	334	22.0	343	4	US-08-997-457-32	Sequence 32, Appli
43	326.5	21.5	1536	4	US-09-413-814-10	Sequence 10, Appli
44	326.5	21.5	3170	4	US-09-036-987A-4	Sequence 4, Appli
45	326.5	21.5	3170	4	US-09-370-700-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-789-609A-2
; Sequence 2, Application US/08789609A
; Patent No. 5827689
; GENERAL INFORMATION:
; APPLICANT: Gentry, Daniel
; APPLICANT: Lonsdale, John
; APPLICANT: Pearson, Stewart
; APPLICANT: Payne, David
; TITLE OF INVENTION: No. 5827689e1 FabD
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/789,609A
; FILING DATE: 24-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/030685
; FILING DATE: 13-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: GM50004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 308 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-789-609A-2

Query Match 36.1%; Score 547.5; DB 2; Length 308;
Best Local Similarity 40.6%; Fred. No. 2.1e-42;
Matches 121; Conservative 54; Mismatches 116; Indels 5; Gaps 4;


```

; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 70
; LENGTH: 2259
; TYPE: PR1
; ORGANISM: Sorangium cellulosum
US-09-413-814-70

Query Match          24.6%; Score 374; DB 4; Length 2259;
Best Local Similarity 34.8%; Pred. No. 3.2e-25;
Matches 101; Conservative 39; Mismatches 135; Indels 16; Gaps 6;

QY 3 KTAFLFAGGGAQYLGNGRDFYDQYPIVKETIDRASQVLGYDLRYLI-----DTEED 53
Db 373 KLAFLFAGGGAQYVPGMGRGLWEAPAFRETDFRCVTLFDRELHQPLCEVWMAEPGSSRSS 432
QY 54 KLNQTYTQPAILLATSVAIYRLLOEKGYOPDMVAGISLGEYSALVASGALDDEDAVALVA 113
Db 433 LLDQTAFTQPALFALEYALAAALFRSGVPELFVAGHSLGELVAAACVAGVFSLEDAVRLVV 492
QY 114 KRGAYMEEAAPADSGRWKVAVLNTPVEVIEEACQKASELGWVTPANTNTPAQIVAGEVVA 173
Db 493 ARGRLM-QALPA-GGAMVSTAAPEADV---AAAVAPHAALVSTAAVNGPEQVVIAGAEKF 547
QY 174 VDRAVELLQEAQAKRILPIKVSQPFHTALLEPASQKLAETLAQVFSDFTCPLVGNTEAA 233
Db 548 VQGIATAAFARGA-RTKPLHVSHAFHSPLMDPMLLEAFRRVTESTYTRFSPIALVSNLSGK 606
QY 234 VMQKEDIAQ-LLTRQVKPEVRFYESIGVMQEAQISNFIETGPGKVLSGFV 282
Db 607 PCTDEVSAPGYVWRHAREAVREADGVKALHAAGAGLFEVGPKPTLLGLV 656

RESULT 6
US-09-335-409-7
; Sequence 7, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCES: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PR1
; ORGANISM: Sorangium cellulosum
US-09-335-409-7

Query Match          24.6%; Score 374; DB 3; Length 2439;
Best Local Similarity 34.8%; Pred. No. 3.6e-25;
Matches 101; Conservative 36; Mismatches 135; Indels 16; Gaps 6;

QY 3 KTAFLFAGGGAQYLGNGRDFYDQYPIVKETIDRASQVLGYDLRYLI-----DTEED 53
Db 553 KLAFLFAGGGAQYVPGMGRGLWEAPAFRETDFRCVTLFDRELHQPLCEVWMAEPGSSRSS 612
QY 54 KLNQTYTQPAILLATSVAIYRLLOEKGYOPDMVAGISLGEYSALVASGALDDEDAVALVA 113
Db 613 LLDQTAFTQPALFALEYALAAALFRSGVPELFVAGHSLGELVAAACVAGVFSLEDAVRLVV 672
QY 114 KRGAYMEEAAPADSGRWKVAVLNTPVEVIEEACQKASELGWVTPANTNTPAQIVAGEVVA 173
Db 673 ARGRLM-QALPA-GGAMVSTAAPEADV---AAAVAPHAALVSTAAVNGPEQVVIAGAEKF 727

```

QY 174 VDRVELLOEAGAKRLIPLKVSPPHTTALLEPASOKLAETLAQVFSDFTCPLVGNTEAA 233
DB 728 VQOIAAFAARGA-RTKPLVSHAFSPMDPMLFAFRVTSVTVRRPSIALVSNLSGK 786
QY 234 VMQKEDIAQ-LLTRQVKEPVRVESIGVMQEAGISNFIIEIGPKVLSGFV 282
DB 787 PCTDEVSAPGYVVRHAREAVRADGVKALHAAGAGLFVEVGPKPTLLGLV 836

RESULT 7
US-09-568-102-7
; Sequence 7, Application US/09568102
; Patent No. 635457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-102-7

Query Match 24.6%; Score 374; DB 4; Length 2439;
Best Local Similarity 34.8%; Pred. No. 3.6e-25;
Matches 101; Conservative 38; Mismatches 135; Indels 16; Gaps 6;
QY 3 KTAFLFAGGGAQVGMGRDFYDQYPIVKETIDRASQVLYDRLYLI-----DTEED 53
DB 553 KLAFLFAGGGAQVGMGRGLWEAWPAFRFTFCVTLPDLHQPLCEVWMAEPGSSRSS 612
QY 54 KLNQTRYTPAILATSVAIYRLLOEKGYOPDMVAGLSGEYSALVAGSALDDEDAVALVA 113
DB 613 LLDQTAFTQALFALEYALALFRSGVGEFELVAGHSLGELVAACVAGVFSLEDAVRLV 672
QY 114 KRGAYVEEAAPADSGKXWAVLNTPEVIEEACQKASELGWVTPANTPAQIVIAGEVVA 173
DB 673 ARGELM-QALPA-GGANVSIAPAEADV---AAAVAPHAALVSIAAVNGPEQVVIAGAEKF 727
QY 174 VDRVELLOEAGAKRLIPLKVSPPHTTALLEPASOKLAETLAQVFSDFTCPLVGNTEAA 233
DB 728 VQOIAAFAARGA-RTKPLVSHAFSPMDPMLFAFRVTSVTVRRPSIALVSNLSGK 786
QY 234 VMQKEDIAQ-LLTRQVKEPVRVESIGVMQEAGISNFIIEIGPKVLSGFV 282
DB 787 PCTDEVSAPGYVVRHAREAVRADGVKALHAAGAGLFVEVGPKPTLLGLV 836

RESULT 8
US-09-567-969-7
; Sequence 7, Application US/09567969
; Patent No. 635457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A

; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-969-7

Query Match 24.6%; Score 374; DB 4; Length 2439;
Best Local Similarity 34.8%; Pred. No. 3.6e-25;
Matches 101; Conservative 38; Mismatches 135; Indels 16; Gaps 6;
QY 3 KTAFLFAGGGAQVGMGRDFYDQYPIVKETIDRASQVLYDRLYLI-----DTEED 53
DB 553 KLAFLFAGGGAQVGMGRGLWEAWPAFRFTFCVTLPDLHQPLCEVWMAEPGSSRSS 612
QY 54 KLNQTRYTPAILATSVAIYRLLOEKGYOPDMVAGLSGEYSALVAGSALDDEDAVALVA 113
DB 613 LLDQTAFTQALFALEYALALFRSGVGEFELVAGHSLGELVAACVAGVFSLEDAVRLV 672
QY 114 KRGAYVEEAAPADSGKXWAVLNTPEVIEEACQKASELGWVTPANTPAQIVIAGEVVA 173
DB 673 ARGELM-QALPA-GGANVSIAPAEADV---AAAVAPHAALVSIAAVNGPEQVVIAGAEKF 727
QY 174 VDRVELLOEAGAKRLIPLKVSPPHTTALLEPASOKLAETLAQVFSDFTCPLVGNTEAA 233
DB 728 VQOIAAFAARGA-RTKPLVSHAFSPMDPMLFAFRVTSVTVRRPSIALVSNLSGK 786
QY 234 VMQKEDIAQ-LLTRQVKEPVRVESIGVMQEAGISNFIIEIGPKVLSGFV 282
DB 787 PCTDEVSAPGYVVRHAREAVRADGVKALHAAGAGLFVEVGPKPTLLGLV 836

RESULT 9
US-09-568-480-7
; Sequence 7, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-480-7

Query Match 24.6%; Score 374; DB 4; Length 2439;
Best Local Similarity 34.8%; Pred. No. 3.6e-25;
Matches 101; Conservative 38; Mismatches 135; Indels 16; Gaps 6;
QY 3 KTAFLFAGGGAQVGMGRDFYDQYPIVKETIDRASQVLYDRLYLI-----DTEED 53
DB 553 KLAFLFAGGGAQVGMGRGLWEAWPAFRFTFCVTLPDLHQPLCEVWMAEPGSSRSS 612
QY 54 KLNQTRYTPAILATSVAIYRLLOEKGYOPDMVAGLSGEYSALVAGSALDDEDAVALVA 113
DB 613 LLDQTAFTQALFALEYALALFRSGVGEFELVAGHSLGELVAACVAGVFSLEDAVRLV 672

QY 114 KRGVMEAPADSGKQWAVLNTPVEVIEEACQKASELGVVTPANYNTPAQIVIAEVA 173
Db 673 ARGSLM-QALPA-GGAMVSIAPADV---AAAVAPHAALVSIAAVNGPQVVIAGAEKF 727
QY 174 VDRAVELLQEGAGKRLIPLKVSQGFHTALLPASPQKLAETLAQVSFSDFTCPVLGVNTEAA 233
Db 728 VQOIAAFAARGA-RTKPLHVSHAFHSLPMDPMLAEFRVTSVYRRPSIALVSNLSGK 786
QY 234 VMQKEDIAQ-LLTRQVKEPVRFYISGVMOEAGISNFIETGPKVLSGFV 282
Db 787 PCTDEVSAPGYWVRHAREAVRFADGVKALHAAGAGLFVEVGPKPTLLGLV 836

RESULT 10

US-09-568-486-7

; Sequence 7, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-486-7

Query Match 24.6%; Score 374; DB 4; Length 2439;
Best Local Similarity 34.8%; Pred. No. 3.6e-25;
Matches 101; Conservative 38; Mismatches 135; Indels 16; Gaps 6;
QY 3 KTAFLPAGCGAQYLGMRDFFDYQYPIVKETIDRASQVGLGYDLRYLI-----DTEED 53
Db 553 KLAFLPAGCGAQVPGMGRLWEAWPAFRETDFRCVTLLFDRELHQLCEVMNAEPGSSRSS 612
QY 54 KINQTRYTOPAILATSVAIYRLLOEKGYQPDVAGLSLGEYSALVSGALDFEDAVALVA 113
Db 613 LLDQTAFTQPALFALEYAALFRSGVPELVAHSLGELVAACVAGVFSLEDAVRLVV 672
QY 114 KRGVMEAPADSGKQWAVLNTPVEVIEEACQKASELGVVTPANYNTPAQIVIAEVA 173
Db 673 ARGSLM-QALPA-GGAMVSIAPADV---AAAVAPHAALVSIAAVNGPQVVIAGAEKF 727
QY 174 VDRAVELLQEGAGKRLIPLKVSQGFHTALLPASPQKLAETLAQVSFSDFTCPVLGVNTEAA 233
Db 728 VQOIAAFAARGA-RTKPLHVSHAFHSLPMDPMLAEFRVTSVYRRPSIALVSNLSGK 786
QY 234 VMQKEDIAQ-LLTRQVKEPVRFYISGVMOEAGISNFIETGPKVLSGFV 282
Db 787 PCTDEVSAPGYWVRHAREAVRFADGVKALHAAGAGLFVEVGPKPTLLGLV 836

RESULT 11

US-09-568-472-7

; Sequence 7, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross

; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-472-7

Query Match 24.6%; Score 374; DB 4; Length 2439;
Best Local Similarity 34.8%; Pred. No. 3.6e-25;
Matches 101; Conservative 38; Mismatches 135; Indels 16; Gaps 6;
QY 3 KTAFLPAGCGAQYLGMRDFFDYQYPIVKETIDRASQVGLGYDLRYLI-----DTEED 53
Db 553 KLAFLPAGCGAQVPGMGRLWEAWPAFRETDFRCVTLLFDRELHQLCEVMNAEPGSSRSS 612
QY 54 KINQTRYTOPAILATSVAIYRLLOEKGYQPDVAGLSLGEYSALVSGALDFEDAVALVA 113
Db 613 LLDQTAFTQPALFALEYAALFRSGVPELVAHSLGELVAACVAGVFSLEDAVRLVV 672
QY 114 KRGVMEAPADSGKQWAVLNTPVEVIEEACQKASELGVVTPANYNTPAQIVIAEVA 173
Db 673 ARGSLM-QALPA-GGAMVSIAPADV---AAAVAPHAALVSIAAVNGPQVVIAGAEKF 727
QY 174 VDRAVELLQEGAGKRLIPLKVSQGFHTALLPASPQKLAETLAQVSFSDFTCPVLGVNTEAA 233
Db 728 VQOIAAFAARGA-RTKPLHVSHAFHSLPMDPMLAEFRVTSVYRRPSIALVSNLSGK 786
QY 234 VMQKEDIAQ-LLTRQVKEPVRFYISGVMOEAGISNFIETGPKVLSGFV 282
Db 787 PCTDEVSAPGYWVRHAREAVRFADGVKALHAAGAGLFVEVGPKPTLLGLV 836

RESULT 12

US-09-567-899-7

; Sequence 7, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-899-7

Query Match 24.6%; Score 374; DB 4; Length 2439;
Best Local Similarity 34.8%; Pred. No. 3.6e-25;
Matches 101; Conservative 38; Mismatches 135; Indels 16; Gaps 6;
QY 3 KTAFLPAGCGAQYLGMRDFFDYQYPIVKETIDRASQVGLGYDLRYLI-----DTEED 53
Db 553 KLAFLPAGCGAQVPGMGRLWEAWPAFRETDFRCVTLLFDRELHQLCEVMNAEPGSSRSS 612

```

QY 54 KINQTRYTPAILATSAIYRLLOEKGYQPKWAGLSLGEYSALVAGSALDFFEDAVLVA 113
Db 613 LLDQTAFTQPALFALEVALAALFRSGVPELVAHSLGELVAACVAGVFSLEDAVLV 672
QY 114 KRGAYMEBAAPADSKWAVLNTPEVIEEACQKASELGVVTPANYNTPAQIVIAGEVVA 173
Db 673 AGRGLX-QALPA-GGAWVSIAPEADV---AAAVAPHAALVSIANVNGPEQVVIAGAEKF 727
QY 174 VRAVELLOEAGAKLILPKVSGPPHTALLPASPQKLAETLAQVFSDFTCPLVGNTEAA 233
Db 728 VQOIAAFAARGA-RTKELHVSHPFSLPMDLEAFRRVTESVYRRPSIALVSNLSGK 786
QY 234 VMQKEDIAQ-LLTROVKPEVRFYESIGVMQAGISNFTIIGPKVLSGFV 282
Db 787 PCTDEVSAPGVWRHAREAVRFAQGVKALHAAGLFFVEVGPKFTLLGLV 836

RESULT 13
US-09-413-814-48
; Sequence 48, Application US/094113814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofile, Gerhard
; APPLICANT: Hofile, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; TITLE OF INVENTION: heteropolyketide compounds
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 48
; LENGTH: 2475
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-48

```

```

Query Match 24.6%; Score 374; DB 4; Length 2475;
Best Local Similarity 33.7%; Pred. No. 3.7e-25;
Matches 99; Conservative 54; Mismatches 119; Indels 22; Gaps 8;

QY 3 KTAFLFAGGAGYLVGMGRDFYDQYF---IVKETIDRASQVLGYDLRYLI--DTEEDKLN 56
Db 1496 RCVFLFPGGAGYPSWARDLVNRCDFALHLDPCLDQLAELLPEDPICILFGDPAERLD 1555
QY 57 QTRYTPAILATSAIYRLLOEKGYQPKWAGLSLGEYSALVAGSALDFFEDAVLVAKRG 116
Db 1556 QTAFTQPLIFSYALARWLGDFIRPDAMIGHLSIGEYVAACLAGLSLSDALLVSERG 1615
QY 117 AYMEBAAPADSKWAVLNTPEVIEEACQKASELGVVTPANYNTPAQIVIAGEVVAADR 176
Db 1616 RUMGSAA---RQAMIAV-PLPWELEERLELLAD-DRISIANVNTAESCVIAGPSEAIR 1670
QY 177 AVLELOEAGAKLILPKVSGPPHTALLPASPQKLAETLAQVFSDFTCPLVGNTE---- 231
Db 1671 CAQRWAAQGL-TCTPLRTSHAFSHAMPEFIVPEPFGHVLARVTFAPPRARMISNLDGKPID 1729
QY 232 -AAVMQKEDIAQLLTROVKPEVRFYESIGVMQAGISNFTIIGPKVLSGFVK 284
Db 1730 SAAVMQPD-----YWRHLINQVRRFEGHSLHLLAETHAWVEVGPRTLUSSPVR 1779

```

```

RESULT 14
US-09-105-537-2
; Sequence 2, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 5215
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-105-537-2

Query Match 24.5%; Score 372.5; DB 4; Length 5215;
Best Local Similarity 33.1%; Pred. No. 1.5e-24;
Matches 104; Conservative 51; Mismatches 132; Indels 27; Gaps 8;

QY 3 KTAFLFAGGAGYLVGMGRDFYDQYFIVKETIDRASQVLGYDLRYLI-----TEEDKL 55
Db 533 RLAVLFSGGAGQRTGGMELVAAHPAFATAFD---AAVAELDPLDRPLAELVAAGDTL 588
QY 56 NQTRYTPAILATSAIYRLLOEKGYQPKWAGLSLGEYSALVAGSALDFFEDAVLVAKR 115
Db 589 DRTVHTQPALFAVEVALHRLVSVGWTPDLLAGHSVGEISAAHVAGVLSLRDAARLVAAR 648
QY 116 GAYMEBAAPADSKWAVLNTPEVIEEACQKASELGVVTPANYNTPAQIVIAGEVVA 175
Db 649 GRIMQ--ALPEGAMVAVEASEEVLPHLAGRERLSL---AAVNGPRAVLAGRAVL 703
QY 176 RAVELLOEAGAKLILPKVSGPPHTALLPASPQKLAETLAQVFSDFTCPLVGNTEAAVM 235
Db 704 DVAELLREQG-RRTKRLSVSHAFHSPMBPMLDDFRVVVEELDFQEPVDDVYVSTVTGLPV 762
QY 236 ---QKEDIAQLLTROVKPEVRFYESIGVMQAGISNFTIIGPKVLSGFVKIDQTAHLA 292
Db 763 TAGQWTD-PEYVWDQVRREVRFLDAVRTLUESGADTFLELPGDGVCSA-----MAADSV 815
QY 293 HVEDQASLVALLEK 306
Db 816 RQERATAVSALRK 829

```

```

RESULT 15
US-08-804-227C-3
; Sequence 3, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kunstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS

```


Search completed: June 11, 2003, 20:06:35
Job time : 35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 20:04:42 ; Search time 309 seconds
(without alignments)
102.238 Million cell updates/sec

Title: US-09-308-397-2

Perfect score: 1518

Sequence: 1 MKTAPLFAQGAQYLMGR.....QTAHLARVEDQASVALLEK 306

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 392085 seqs, 103240269 residues

Total number of hits satisfying chosen parameters: 392085

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/prodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/prodata/1/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1514	99.7	306	10	US-09-815-242-13580
2	937	61.7	313	10	Sequence 13580, A
3	643.5	42.4	309	10	Sequence 10920, A
4	636	41.9	312	10	Sequence 10125, A
5	586.5	36.7	302	10	Sequence 10975, A
6	556.5	36.7	311	10	Sequence 5460, Ap
7	556.5	36.7	312	10	Sequence 12130, A
8	416.5	27.4	309	10	Sequence 11894, A
9	414.5	27.3	1402	9	Sequence 11290, A
10	376.5	24.8	4150	9	Sequence 166, App
11	374	24.6	2439	9	Sequence 2, Appli
12	372.5	24.5	5215	9	Sequence 7, Appli
13	372.5	24.5	5215	9	Sequence 2, Appli
14	372.5	24.5	5215	9	Sequence 2, Appli
15	372.5	24.5	5215	10	Sequence 2, Appli
16	371	24.4	1616	9	Sequence 262, App
17	356	23.5	7257	9	Sequence 5, Appli
18	351	23.1	1421	9	Sequence 2, Appli
19	346.5	23.0	2756	9	Sequence 7, Appli

20	348	22.9	395	9	US-10-074-045-46	Sequence 46, Appl
21	340.5	22.4	426	9	US-10-124-800-10	Sequence 10, Appl
22	340.5	22.4	2910	9	US-10-124-800-2	Sequence 2, Appli
23	334	22.0	343	9	US-09-735-556-32	Sequence 32, Appl
24	324	21.3	4613	9	US-09-860-846-31	Sequence 31, Appl
25	324	21.3	4613	9	US-09-988-384B-31	Sequence 31, Appl
26	324	21.3	4613	9	US-09-836-821-31	Sequence 31, Appl
27	324	21.3	4613	10	US-09-861-289-31	Sequence 31, Appl
28	324	21.3	11877	9	US-09-860-846-6	Sequence 6, Appli
29	324	21.3	11877	9	US-09-836-821-6	Sequence 6, Appli
30	324	21.3	11877	10	US-09-861-289-6	Sequence 6, Appli
31	324	21.3	12199	9	US-09-988-384B-6	Sequence 31, Appl
32	306	20.2	328	9	US-09-735-556-31	Sequence 4, Appli
33	275.5	18.1	1832	9	US-10-014-717-4	Sequence 6, Appli
34	274	18.1	3798	9	US-10-014-717-6	Sequence 46, Appl
35	254	16.7	2563	9	US-09-836-705-46	Sequence 4, Appli
36	253.5	16.7	3519	9	US-09-808-880-4	Sequence 3, Appli
37	246	16.2	3816	9	US-09-808-880-3	Sequence 261, App
38	235.5	15.5	1827	9	US-09-712-363-261	Sequence 34, Appl
39	235	15.5	345	9	US-09-735-556-34	Sequence 6666, Ap
40	221	14.6	1610	9	US-09-738-626-6666	Sequence 35, Appl
41	217.5	14.3	1562	9	US-09-860-846-35	Sequence 35, Appl
42	217.5	14.3	1562	9	US-09-988-384B-35	Sequence 35, Appl
43	217.5	14.3	1562	9	US-09-836-821-35	Sequence 35, Appl
44	217.5	14.3	1562	10	US-09-861-289-35	Sequence 35, Appl
45	212	14.0	1346	9	US-09-860-846-37	Sequence 37, Appl

ALIGNMENTS

RESULT 1
US-09-815-242-13580
; Sequence 13580, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13580
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13580

Query Match 99.7%; Score 1514; DB 10; Length 306;
Best Local Similarity 99.7%; Pred. No. 5.2e-124;

Matches 305; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKTAFAGQAQYGLMGRDFDYQYPIVKETIDRASQVLGYDLRYLIDTEEDKLNQTRY 60
DB 1 MTKTAFAGQAQYGLMGRDFDYQYPIVKETIDRASQVLGYDLRYLIDTEEDKLNQTRY 60

QY 61 TOPAILATSVAIYRLLOEKGYOPDMVAGLSLGEYSALVASCALDPEDAVAVAKGAYME 120
DB 61 TOPAILATSVAIYRLLOEKGYOPDMVAGLSLGEYSALVASCALDPEDAVAVAKGAYME 120

QY 121 EAAPADSGMWAVLNTPVEVIEEACQKASELGVVTPANYNTPAQIVIAGEVAVDRAVEL 180
DB 121 EAAPADSGMWAVLNTPVEVIEEACQKASELGVVTPANYNTPAQIVIAGEVAVDRAVEL 180

QY 181 LOERAKELIPLKVSQPHHTALLEPASOKLAETTLAQVSFSDFTCPLVGNTEAAVMQKEDI 240
DB 181 LOERAKELIPLKVSQPHHTALLEPASOKLAETTLAQVSFSDFTCPLVGNTEAAVMQKEDI 240

QY 241 AQLLTRQVKEPVRVYESIGVMOEAGISNFIETGPKVLSGFVKKIDOTAHLAHVEDQASL 300
DB 241 AQLLTRQVKEPVRVYESIGVMOEAGISNFIETGPKVLSGFVKKIDOTAHLAHVEDQASL 300

QY 301 VALLEK 306
DB 301 VALLEK 306

RESULT 2

US-09-815-242-10920

; Sequence 10920, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 10920

; LENGTH: 313

; TYPE: PRN

; ORGANISM: Enterococcus faecalis

US-09-815-242-10920

Query Match 61.7%; Score 937; DB 10; Length 313;

Best Local Similarity 61.9%; Pred. No. 9,3e-74;

Matches 187; Conservative 40; Mismatches 75; Indels 0; Gaps 0;

QY 3 KTAFLPAGQAQYGLMGRDFDYQYPIVKETIDRASQVLGYDLRYLIDTEEDKLNQTRY 62

Db 6 KTAFLPAGQAQYGLMGRDFDYQYPIVKETIDRASQVLGYDLRYLIDTEEDKLNQTRY 65

QY 63 PAILATSVAIYRLLOEKGYOPDMVAGLSLGEYSALVASCALDPEDAVAVAKGAYMEEA 122

Db 66 PAILTSVAVAFYRLLOEKGYOPDMVAGLSLGEYSALVASCALDPEDAVAVAKGAYMEEA 125

QY 123 APADSGMWAVLNTPVEVIEEACQKASELGVVTPANYNTPAQIVIAGEVAVDRAVELIQ 182

Db 126 APOGTGKMWAVNAREVIEEACQKASELGVVTPANYNTPAQIVIAGEVAVDRAVELIQ 185

QY 183 ERAGAKELIPLKVSQPHHTALLEPASOKLAETTLAQVSFSDFTCPLVGNTEAAVMQKEDIAQ 242

Db 186 EAGVEMIFLNVSGPHHTALLEPASOKLAETTLAQVSFSDFTCPLVGNTEAAVMQKEDIAQ 245

QY 243 LLTRQVKEPVRVYESIGVMOEAGISNFIETGPKVLSGFVKKIDOTAHLAHVEDQASLVA 302

Db 246 LLEKQVMSAVRPFESIGVMOEAGISNFIETGPKVLSGFVKKIDOTAHLAHVEDQASLVA 305

QY 303 LL 304

Db 306 TL 307

RESULT 3

US-09-815-242-10125

; Sequence 10125, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 10125

; LENGTH: 309

; TYPE: PRN

; ORGANISM: Escherichia coli

US-09-815-242-10125

Query Match 42.4%; Score 643.5; DB 10; Length 309;

Best Local Similarity 45.5%; Pred. No. 3,3e-48;

Matches 141; Conservative 60; Mismatches 102; Indels 7; Gaps 4;

QY 1 MTKTAFAGQAQYGLMGRDFDYQYPIVKETIDRASQVLGYDLRYLIDTEEDKLNQ 57

Db 1 MTQFAVFFGQSQVTVGLADMAASYPVIEETFAEASALGYDLWAL--TQGGPAEELNK 58

QY 58 TRYTPAILATSVAIYRLLOEKGYO-PDMVAGLSLGEYSALVASCALDPEDAVAVAKRG 116

```
Db 59 TWQTQALLTASVALYRWQGGKAPAMMAGHSLGEYSALVCAGVIDFADAVRLVEMRG 118
QY 117 AYMEBAAPADSGKMVAVLNTPVEVIEACQKASGLGVVTPANTPAQIVVAGEVAVDR 176
Db 119 KFMQEAPEPFGMAAIIIGLDASIAKACBAEAGQVSPVFNFSPOGVVIAGKEAVER 178
QY 177 AVELLOBAGAKRLPLKVGSPFFHTALLEPASOKLAETLAQVSPSDFCTCLVGNTEAAVMQ 236
Db 179 AGAACKAAGAKRALPLPVSPSHCALMKPAADKLAVELAKITNAPTVPVNNVVKCET 238
QY 237 KED-IAQLLTRQVKPEVRFYESIGVMQEAAGISNFIETGPGKVLSGFVKKIDQTAHLAHVE 295
Db 239 NGDAIRDALVQLYNPQWTKSVYMAAQGVHELYEVPKGLTGLTKRIVDTLTASALN 298
QY 296 DQASLVALLE 305
Db 299 EPSAMAAALE 308

RESULT 4
US-09-815-242-10975
; Sequence 10975, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10975
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-10975

Query Match 41.9%; Score 636; DB 10; Length 312;
Best Local Similarity 46.0%; Pred. No. 1.5e-47;
Matches 142; Conservative 54; Mismatches 109; Indels 4; Gaps 3;

QY 1 MTKTAPLAFAGGAQVLMGRDFYQYPIVKETIDRASQVLYGDIYRILDT-EDKLNQTR 59
Db 1 MKKTAMVFPFGGSGTQVGMGLADLATEYPIVETPKQASDALGYDLYWLVQQGPAEELNKTW 60
QY 60 YTOPAILATSVAIYRLLEKGYQ-PMVAGLSLGEYSALVASCALDFEDAVALKRGA 117
Db 61 QTQPALLAASVAITRVWKEKFPQKPENWAGHSLGEYSALVCAGVLDFQDAIKLVELRGK 120
QY 118 YMEBAAPADSGKMVAVLNTPVEVIEACQKASELGVVTPANTPAQIVVAGEVAVDR 177
```

```
Db 121 LMQQAVPSTGMVYAIIGLDNEAIINACKQAECEVSVAVNFNSPGOVVIACAKAVERA 180
QY 178 VELLOBAGAKRLPLKVGSPFFHTALLEPASOKLAETLAQVSPSDFCTCLVGNTEA-AVMQ 236
Db 181 AALCKEAGAKRALPLAVSPSHCALMKPAEGLAVTLENIQINTPTISVLNVDVKAETE 240
QY 237 KEDIAQLLTRQVKPEVRFYESIGVMQEAAGISNFIETGPGKVLSGFVKKIDQTAHLAHVED 296
Db 241 GTEITALTVRQYSPVRMTETVVKVQAQGVLVLAIEVGGKVLNGLTKRIVGDLQALSVND 300
QY 297 QASLVALLE 305
Db 301 VASFNAAVEE 309

RESULT 5
US-09-815-242-5460
; Sequence 5460, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5460
; LENGTH: 302
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5460

Query Match 36.7%; Score 556.5; DB 10; Length 302;
Best Local Similarity 40.9%; Pred. No. 1.2e-40;
Matches 122; Conservative 54; Mismatches 117; Indels 5; Gaps 4;

QY 1 MTKTAPLAFAGGAQVLMGRDFYQYPIVKETIDRASQVLYGDIYRILDTBED-KLNQTR 59
Db 1 MSKTAIIFPGGAQVKGVAQDLYNNNDQATEILTSAAANTLDFDILETMTFTDGDGKLGETE 60
QY 60 YTOPAILATSVAIYRLLEKGYQPMVAGLSLGEYSALVASCALDFEDAVALKRGA 119
Db 61 NTQPALLTSSALLAAL--KNLNPFTMGHSLGEYSYSSVAADVLSEFEDAVKIVRKGQIM 118
QY 120 EBAAPADSGKMVAVLNTPVEVIEACQK-ASELGVVTPANTPAQIVVAGEVAVDR 178
Db 119 AQAPFTGVGSMARVLGLDFDKVDEICKSLSSDDKIIIEPANINCPQIVVSGHKALIDELV 178
QY 179 ELLQBAGAKRLPLKVGSPFFHTALLEPASOKLAETLAQVSPSDFCTCLVGNTEA-AVMQ 237
```

```
Db 179 EKGSLGAKRWPLAVSGFFHSSLMKVIEEDFSSYINQFEWHDAKFPVQVNVNAQGETDK 238
QY 238 EDIAQLLTRQVKPEVRFYES:GVMQEGAGISNFIEIGPGKVLGSGFVKXIDQTAHLAHVE 295
Db 239 EVIKSNMVKQLYSPVQFINSTEWLIDQGVDFHIEIGPGKVLGSLGIKKINRDVKLTSLQ 296

RESULT 6
US-09-815-242-12130
; Sequence 12130, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12130
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12130

Query Match 36.7%; Score 556.5; DB 10; Length 311;
Best Local Similarity 40.9%; Pred. No. 1.2e-40;
Matches 122; Conservative 54; Mismatches 117; Indels 5; Gaps 4;

QY 1 MTKTAPLFGQGAQYLGMRDIFYDQYPIVKETIDRASQVLGYDLRYLIDTEED-KLNQTR 59
Db 4 MSKTAIFPGQGAQYGVMAQQLYNNNDQATEILTSAAANTLDFDILEMTFTDEDGKLGETE 63

QY 60 VTQPAILATSVAIYELLQKGYQPDWAGLSLGEYSALVASGALDFEDAVALKVAKRGAYM 119
Db 64 NTQPAILLTHSGALLAAL--KMLNPDFTWGHSLGETSSLSVAADVUSFEDAVKIVKRGQLM 121

QY 120 EEAPADSGKMWAVLNTPEVEIEACQK-ASELGWVTPANTYNTPAQIVTAGEVAVDRAV 178
Db 122 AQAPFTGVGMAAVLGLDFDKVDKICKSLSSDDKIIPEANINCPGQIVWSGHKALIDELV 181

QY 179 ELLQZAGAKRLIPLKVSQGFHTALLEPASQKLAETLAQVSFSDFTCLVGNTEA-AVMQK 237
Db 182 EKGSLGAKRWPLAVSGFFHSSLMKVIEEDFSSYINQFEWHDAKFPVQVNVNAQGETDK 241

QY 238 EDIAQLLTRQVKPEVRFYES:GVMQEGAGISNFIEIGPGKVLGSGFVKXIDQTAHLAHVE 295
Db 242 EVIKSNMVKQLYSPVQFINSTEWLIDQGVDFHIEIGPGKVLGSLGIKKINRDVKLTSLQ 299

RESULT 7
```

```
US-09-815-242-11894
; Sequence 11894, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11894
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11894

Query Match 36.7%; Score 556.5; DB 10; Length 312;
Best Local Similarity 44.5%; Pred. No. 1.2e-40;
Matches 126; Conservative 51; Mismatches 103; Indels 3; Gaps 3;

QY 5 AFLFAGQGAQYLGMRDIFYDQYPIVKETIDRASQVLGYDLRYLIDT-BEDKLNQTRYTOP 63
Db 6 AFVFPQGSQSLGMLAEALGAQQAQLVRDTFAEASEALGYDLWALVQNGPEERLNTDKTOP 65

QY 64 AILATSVATYRL-LOEKGYPQDMVAGLSLGEYSALVASGALDFEDAVALKVAKRGAYMEEA 122
Db 66 AILTVSIALWRLWLAEAGNAPAFVAGHSLGEYSALVAESLAFADAVKLVBERGQLMQQA 125

QY 123 APADSGKMWAVLNTPEVEIEACQKASBELGVVTPANTYNTPAQIVTAGEVAVDRAVELLQ 182
Db 126 VPAGQGMMAAILGLEDDADVLAAACAEAAQGVVSAVNFNAPGQVVIAGAAAVERAIEACK 185

QY 183 EAGAKELIPLKVSQGFHTALLEPASQKLAETLAQVSFSDFTCLVGNTEAAVMQKED-IA 241
Db 186 ARGAKRAVALPVSVSHCELMPAAEQFAASVESLQWQAPKISLVQNVSAAPADLDILR 245

QY 242 QLLTRQVKPEVRFYESIGVMQEGAGISNFIEIGPGKVLGSGFVKK 284
Db 246 RDLAQLYSPVAVWVESIQLLAEKGYTELVECGPGKVLGNRR 288

RESULT 8
US-09-815-242-11290
; Sequence 11290, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
```



```
; LENGTH: 4150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
; OTHER INFORMATION: Oleandolide PKS
US-09-808-880-2

Query Match      24.8%; Score 376.5; DB 9; Length 4150;
Best Local Similarity 33.2%; Pred. No. 2.1e-23;
Matches 99; Conservative 52; Mismatches 122; Indels 25; Gaps 5;

QY 5 AFLEAGGAGYLGMRDGYDQYPIVKETIDRASQVIGYDLRYLIDTEEDK-----54
Db 563 APLFGGQSQQPMGRKRLHQVFFGRDALDEVCAELDTHLGRLLGPPAGPPLRVMFAER 622
QY 55 -----LNQTRYTOPALATSVAIYRLLOEKGYQPMVAGLSLGEYSALVASGALDPED 107
Db 623 GTASHALLSETHYTQALFALETALEFRLLVQWGLKPDHLAGHSVGEIAAAHAAGILDSD 682
QY 108 AVALYAKGAYMEEAAPADSGKMWAVLNTPEVIEEACQKASELGVVTPANYNTPAQIVI 167
Db 683 AELVATRGALMSL--PGGVMLSVQAPSEVAPLLLGREAHVGL--AAVNGPDPAVVV 737
QY 168 AGEVAVDRAVELLOEAGAKRLIPLKVSFPHTALLEPASQKLAETLAQVSFSDFTCPVLV 227
Db 738 SCERGHVAAIEQILDRGRKSRV-LRVSHAFHSPLMPEVLEFAEAVAGLTFRAPTTPLV 796
QY 228 GNTAAVNMKEDIA--QLLTQVKEPVRFYESIGVMQEAGISNFIEIGPGKVLGSGFVK 283
Db 797 SNLTGAPVDDRTMATPAYVWRHREAVRFGDGRALGKLGTSFLEVGPDGVLTMAR 854

RESULT 11
US-10-014-717-7
; Sequence 7, Application US/10014717
; Publication No. US20020192778A1
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/10/014,717
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US/09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-10-014-717-7

Query Match      24.6%; Score 374; DB 9; Length 2439;
Best Local Similarity 34.8%; Pred. No. 1.7e-23;
Matches 101; Conservative 38; Mismatches 135; Indels 16; Gaps 6;

QY 3 KTAFLPAGGAGYLGMRDGYDQYPIVKETIDRASQVIGYDLRYLI-----DTEED 53
Db 553 KLAFLPAGGAGYVPMGRGLMEAWFAFRETDFRCVTFLFRELHOPCEVWMAEPGSSRSS 612
QY 54 KLNQTRYTOPALATSVAIYRLLOEKGYQPMVAGLSLGEYSALVASGALDPEDAVLVA 113
Db 613 LLDQTAFTQPALFALEYALALFRSMGVEPELVAGHSIGELVAACVAGVPSLEDAVLVV 672
QY 114 KEGAYMEEAAPADSGKMWAVLNTPEVIEEACQKASELGVVTPANYNTPAQIIVAGEVVA 173
Db 673 ARGRLM-QALPA-GGAMYSIAAPEADV---AAAVAPHAALYSIAAVNGPEQVVIAGAEKF 727
```

```
QY 174 VDRAVELLOEAGAKRLIPLKVSFPHTALLEPASQKLAETLAQVSFSDFTCPVLGNTEAA 233
Db 728 VQIATAAFAARGA-RTKPLHVSFAHSPMLDPMLEAFRRVTESVTYRRPSIALVSNLSGK 786
QY 234 VMQKEDIAQ-LLTQVKEPVRFYESIGVMQEAGISNFIEIGPGKVLGSGFV 282
Db 787 PCTDEVSAFGYVWRHREAVRFGDGVKALHAAGAGLFVEVGPKPITLLGLV 836

RESULT 12
US-09-860-846-2
; Sequence 2, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 5215
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-2

Query Match      24.5%; Score 372.5; DB 9; Length 5215;
Best Local Similarity 33.1%; Pred. No. 6.4e-23;
Matches 104; Conservative 51; Mismatches 132; Indels 27; Gaps 8;

QY 3 KTAFLPAGGAGYLGMRDGYDQYPIVKETIDRASQVIGYDLRYLI-----TEEDKL 55
Db 533 RLAVLFGGQAGQRTGTGMELVAAHPAFATFD----AVAAELDPLDRPLAELVAGDTL 588
QY 56 NQTRYTOPALATSVAIYRLLOEKGYQPMVAGLSLGEYSALVASGALDPEDAVLVAKR 115
Db 589 DRVHTQPALFAVEVALHRLVESWGTTPDLLAGHSVGEISAAHVAGVLSLRDAARLVAR 648
QY 116 GAYMEEAAPADSGKMWAVLNTPEVIEEACQKASELGVVTPANYNTPAQIIVAGEVAVD 175
Db 649 GRLMQ--ALPEGGAMVAVEASEEVLPLAGRELSL---AAVNGPRAVLGAEAVL 703
QY 176 RAVELLOEAGAKRLIPLKVSFPHTALLEPASQKLAETLAQVSFSDFTCPVLGNTEAAVM 235
Db 704 DVAELLREQG-RTKRLSVSHAFHSPMLPEMDDFRRVVVEELDFOEPRVDVVSTVTGLPV 762
QY 236 ---QKEDIAQLLTQVKEPVRFYESIGVMQEAGISNFIEIGPGKVLGSGFVKIDQTAHLA 292
Db 763 TAGQWTD-PEYVWDQVRREPRLDAVRTLEESGADTFLELGPDGVCSA-----MAADSV 815
QY 293 HVEDQASLVALLEK 306
Db 816 RQERATAVSALRK 829

RESULT 13
US-09-988-384B-2
; Sequence 2, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
```

BEST AVAILABLE COPY

```

; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 2
; LENGTH: 5215
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-2

Query Match      24.5%; Score 372.5; DB 9; Length 5215;
Best Local Similarity 33.1%; Pred. No. 6.4e-23;
Matches 104; Conservative 51; Mismatches 132; Indels 27; Gaps 8;

QY 3 KTAFLFAGGQAQYLGMRDGYDQYPIVKETIDRASQVGLGYDLRYLID-----TEEDKL 55
DB 533 RLAVLFSGGQAQRTGGMELYAHPAFATFD----AAVAELDPLLDRPLAELVAAGDTL 588

QY 56 NOTRYTOPAILATSVAIYRLLOEKGYOPDMVAGLSGEYSALVAGSALDGEDAVALVAKR 115
DB 589 DRTVHTQPALFAVEALHRLVSVGWTPDLLAGHSVGEISAAHVAGVLSLRDAARLVAAR 648

QY 116 GAYMEEAAPADSGKMWAVLNTPEVEIEACOKASELGVVTPANYNTPAQIVIAGEVVAVD 175
DB 649 GRLMQ--ALPEGGMAMVAEASEEVLPHLAGRERELSL---AAVNGPRAVVLAGAERAVL 703

QY 176 RAVELLQAGAKRLIPLKVSQGFHTALLEPASQKLAETLAQVSPDFTCPVLGNTEAAVM 235
DB 704 DVAELLREQG--RRTKRLSVSHAFSPMLPEMLDDFRVVEELDFQEPVVDVSTVTGLPV 762

QY 236 ---QKEDIAQLLTROVKEPVRFYESIGVMOBAGISNFIETGPGKVLGSGFVKKIDQTAHLA 292
DB 763 TAGQWTD--PEYWDQVRRPVRFLDAVRTLEESGADTFLELGPDCGYCSA-----MAADSV 815

QY 293 HVEDQASLVALLEK 306
DB 816 RDOEAATAVSALRK 829

RESULT 14
US-09-836-821-2
; Sequence 2, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 2001-04-17
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 5215
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-836-821-2

Query Match      24.5%; Score 372.5; DB 9; Length 5215;
Best Local Similarity 33.1%; Pred. No. 6.4e-23;
Matches 104; Conservative 51; Mismatches 132; Indels 27; Gaps 8;

QY 3 KTAFLFAGGQAQYLGMRDGYDQYPIVKETIDRASQVGLGYDLRYLID-----TEEDKL 55
DB 533 RLAVLFSGGQAQRTGGMELYAHPAFATFD----AAVAELDPLLDRPLAELVAAGDTL 588

QY 56 NOTRYTOPAILATSVAIYRLLOEKGYOPDMVAGLSGEYSALVAGSALDGEDAVALVAKR 115
DB 589 DRTVHTQPALFAVEALHRLVSVGWTPDLLAGHSVGEISAAHVAGVLSLRDAARLVAAR 648

QY 116 GAYMEEAAPADSGKMWAVLNTPEVEIEACOKASELGVVTPANYNTPAQIVIAGEVVAVD 175
DB 649 GRLMQ--ALPEGGMAMVAEASEEVLPHLAGRERELSL---AAVNGPRAVVLAGAERAVL 703

QY 176 RAVELLQAGAKRLIPLKVSQGFHTALLEPASQKLAETLAQVSPDFTCPVLGNTEAAVM 235
DB 704 DVAELLREQG--RRTKRLSVSHAFSPMLPEMLDDFRVVEELDFQEPVVDVSTVTGLPV 762

QY 236 ---QKEDIAQLLTROVKEPVRFYESIGVMOBAGISNFIETGPGKVLGSGFVKKIDQTAHLA 292
DB 763 TAGQWTD--PEYWDQVRRPVRFLDAVRTLEESGADTFLELGPDCGYCSA-----MAADSV 815

QY 293 HVEDQASLVALLEK 306
DB 816 RDOEAATAVSALRK 829

```

```

; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 2
; LENGTH: 5215
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-2

Query Match      24.5%; Score 372.5; DB 9; Length 5215;
Best Local Similarity 33.1%; Pred. No. 6.4e-23;
Matches 104; Conservative 51; Mismatches 132; Indels 27; Gaps 8;

QY 3 KTAFLFAGGQAQYLGMRDGYDQYPIVKETIDRASQVGLGYDLRYLID-----TEEDKL 55
DB 533 RLAVLFSGGQAQRTGGMELYAHPAFATFD----AAVAELDPLLDRPLAELVAAGDTL 588

QY 56 NOTRYTOPAILATSVAIYRLLOEKGYOPDMVAGLSGEYSALVAGSALDGEDAVALVAKR 115
DB 589 DRTVHTQPALFAVEALHRLVSVGWTPDLLAGHSVGEISAAHVAGVLSLRDAARLVAAR 648

QY 116 GAYMEEAAPADSGKMWAVLNTPEVEIEACOKASELGVVTPANYNTPAQIVIAGEVVAVD 175
DB 649 GRLMQ--ALPEGGMAMVAEASEEVLPHLAGRERELSL---AAVNGPRAVVLAGAERAVL 703

QY 176 RAVELLQAGAKRLIPLKVSQGFHTALLEPASQKLAETLAQVSPDFTCPVLGNTEAAVM 235
DB 704 DVAELLREQG--RRTKRLSVSHAFSPMLPEMLDDFRVVEELDFQEPVVDVSTVTGLPV 762

QY 236 ---QKEDIAQLLTROVKEPVRFYESIGVMOBAGISNFIETGPGKVLGSGFVKKIDQTAHLA 292
DB 763 TAGQWTD--PEYWDQVRRPVRFLDAVRTLEESGADTFLELGPDCGYCSA-----MAADSV 815

QY 293 HVEDQASLVALLEK 306
DB 816 RDOEAATAVSALRK 829

RESULT 14
US-09-836-821-2
; Sequence 2, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 2001-04-17
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 5215
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-836-821-2

Query Match      24.5%; Score 372.5; DB 9; Length 5215;
Best Local Similarity 33.1%; Pred. No. 6.4e-23;
Matches 104; Conservative 51; Mismatches 132; Indels 27; Gaps 8;

QY 3 KTAFLFAGGQAQYLGMRDGYDQYPIVKETIDRASQVGLGYDLRYLID-----TEEDKL 55
DB 533 RLAVLFSGGQAQRTGGMELYAHPAFATFD----AAVAELDPLLDRPLAELVAAGDTL 588

QY 56 NOTRYTOPAILATSVAIYRLLOEKGYOPDMVAGLSGEYSALVAGSALDGEDAVALVAKR 115
DB 589 DRTVHTQPALFAVEALHRLVSVGWTPDLLAGHSVGEISAAHVAGVLSLRDAARLVAAR 648

QY 116 GAYMEEAAPADSGKMWAVLNTPEVEIEACOKASELGVVTPANYNTPAQIVIAGEVVAVD 175
DB 649 GRLMQ--ALPEGGMAMVAEASEEVLPHLAGRERELSL---AAVNGPRAVVLAGAERAVL 703

QY 176 RAVELLQAGAKRLIPLKVSQGFHTALLEPASQKLAETLAQVSPDFTCPVLGNTEAAVM 235
DB 704 DVAELLREQG--RRTKRLSVSHAFSPMLPEMLDDFRVVEELDFQEPVVDVSTVTGLPV 762

QY 236 ---QKEDIAQLLTROVKEPVRFYESIGVMOBAGISNFIETGPGKVLGSGFVKKIDQTAHLA 292
DB 763 TAGQWTD--PEYWDQVRRPVRFLDAVRTLEESGADTFLELGPDCGYCSA-----MAADSV 815

QY 293 HVEDQASLVALLEK 306
DB 816 RDOEAATAVSALRK 829

```

```

; Sequence 2, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 5215
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-861-289-2

Query Match      24.5%; Score 372.5; DB 10; Length 5215;
Best Local Similarity 33.1%; Pred. No. 6.4e-23;
Matches 104; Conservative 51; Mismatches 132; Indels 27; Gaps 8;

QY 3 KTAFLFAGGQAQYLGMRDGYDQYPIVKETIDRASQVGLGYDLRYLID-----TEEDKL 55
DB 533 RLAVLFSGGQAQRTGGMELYAHPAFATFD----AAVAELDPLLDRPLAELVAAGDTL 588

QY 56 NOTRYTOPAILATSVAIYRLLOEKGYOPDMVAGLSGEYSALVAGSALDGEDAVALVAKR 115
DB 589 DRTVHTQPALFAVEALHRLVSVGWTPDLLAGHSVGEISAAHVAGVLSLRDAARLVAAR 648

QY 116 GAYMEEAAPADSGKMWAVLNTPEVEIEACOKASELGVVTPANYNTPAQIVIAGEVVAVD 175
DB 649 GRLMQ--ALPEGGMAMVAEASEEVLPHLAGRERELSL---AAVNGPRAVVLAGAERAVL 703

QY 176 RAVELLQAGAKRLIPLKVSQGFHTALLEPASQKLAETLAQVSPDFTCPVLGNTEAAVM 235
DB 704 DVAELLREQG--RRTKRLSVSHAFSPMLPEMLDDFRVVEELDFQEPVVDVSTVTGLPV 762

QY 236 ---QKEDIAQLLTROVKEPVRFYESIGVMOBAGISNFIETGPGKVLGSGFVKKIDQTAHLA 292
DB 763 TAGQWTD--PEYWDQVRRPVRFLDAVRTLEESGADTFLELGPDCGYCSA-----MAADSV 815

QY 293 HVEDQASLVALLEK 306
DB 816 RDOEAATAVSALRK 829

```

```

; Sequence 2, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 5215
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-861-289-2

Query Match      24.5%; Score 372.5; DB 10; Length 5215;
Best Local Similarity 33.1%; Pred. No. 6.4e-23;
Matches 104; Conservative 51; Mismatches 132; Indels 27; Gaps 8;

QY 3 KTAFLFAGGQAQYLGMRDGYDQYPIVKETIDRASQVGLGYDLRYLID-----TEEDKL 55
DB 533 RLAVLFSGGQAQRTGGMELYAHPAFATFD----AAVAELDPLLDRPLAELVAAGDTL 588

QY 56 NOTRYTOPAILATSVAIYRLLOEKGYOPDMVAGLSGEYSALVAGSALDGEDAVALVAKR 115
DB 589 DRTVHTQPALFAVEALHRLVSVGWTPDLLAGHSVGEISAAHVAGVLSLRDAARLVAAR 648

QY 116 GAYMEEAAPADSGKMWAVLNTPEVEIEACOKASELGVVTPANYNTPAQIVIAGEVVAVD 175
DB 649 GRLMQ--ALPEGGMAMVAEASEEVLPHLAGRERELSL---AAVNGPRAVVLAGAERAVL 703

QY 176 RAVELLQAGAKRLIPLKVSQGFHTALLEPASQKLAETLAQVSPDFTCPVLGNTEAAVM 235
DB 704 DVAELLREQG--RRTKRLSVSHAFSPMLPEMLDDFRVVEELDFQEPVVDVSTVTGLPV 762

QY 236 ---QKEDIAQLLTROVKEPVRFYESIGVMOBAGISNFIETGPGKVLGSGFVKKIDQTAHLA 292
DB 763 TAGQWTD--PEYWDQVRRPVRFLDAVRTLEESGADTFLELGPDCGYCSA-----MAADSV 815

QY 293 HVEDQASLVALLEK 306
DB 816 RDOEAATAVSALRK 829

```


Search completed: June 11, 2003, 20:17:37
Job time : 310 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 19:47:12 ; Search time 44 Seconds
(without alignments)
668.571 Million cell updates/sec

Title: US-09-308-397-2
Perfect score: 1518
Sequence: 1 MTKTAFIFAGGQAQYLGMR.....QTAHLAHVEDQASLVALLEK 306

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1518	100.0	306	2	F95048 malonyl CoA-acyl c
2	1515	99.8	306	2	D97919 [acyl-carrier-prot
3	1048	69.0	308	2	E86721 hypothetical prote
4	713.5	47.0	308	2	G97338 malonyl CoA-acyl c
5	669.5	44.1	317	2	H69620 [acyl-carrier-prot
6	657.5	43.3	312	2	G82128 [acyl-carrier-prot
7	654	43.1	313	2	D83961 malonyl CoA-acyl c
8	643.5	42.4	309	1	E44856 [acyl-carrier-prot
9	640.5	42.2	309	2	F90812 malonyl-CoA-[acyl-
10	636	41.9	312	2	B85672 [acyl-carrier-prot
11	636	41.9	312	2	E64051 [acyl-carrier-prot
12	628.5	41.4	309	2	A10194 malonyl CoA-acyl c
13	612.5	40.3	309	2	AC0642 [acyl-carrier-prot
14	611.5	40.3	307	2	T44433 [acyl-carrier-prot
15	594.5	39.2	326	2	F70376 [acyl-carrier-prot
16	588.5	38.8	313	2	AH1300 malonyl CoA-acyl c
17	586	38.6	313	2	AH1672 malonyl CoA-acyl c
18	570	37.5	308	2	B81026 [acyl-carrier-prot
19	566	37.3	308	2	B81026 [acyl-carrier-prot
20	556.5	36.7	312	2	T12019 malonyl CoA-acyl c
21	551.5	36.3	308	2	E89896 [acyl-carrier-prot
22	533.5	35.1	317	2	E82776 [acyl-carrier-prot
23	531	35.0	308	2	A86528 malonyl acyl carri
24	529	34.8	305	2	A75334 [acyl-carrier-prot
25	528	34.8	308	2	E72096 [acyl-carrier-prot
26	526	34.7	293	2	G72334 [acyl-carrier-prot
27	492	32.4	331	2	C97493 malonyl-CoA-acyl c
28	492	32.4	331	2	AE2711 malonyl-CoA:acyl c
29	485	31.9	314	2	AH3436 [acyl-carrier-prot

30	479	31.6	292	2	AH1836 malonyl coenzyme A
31	479	31.6	314	2	H87456 malonyl CoA-acyl c
32	468	30.6	314	2	D97839 hypothetical prote
33	464.5	30.6	308	2	F81895 [acyl-carrier-prot
34	461.5	30.4	306	2	D81428 [acyl-carrier-prot
35	459	30.2	293	2	S75355 probable [acyl-car
36	457.5	30.1	400	2	T44805 malonyl-CoA transa
37	451.5	29.7	308	2	G71538 [acyl-carrier-prot
38	445	29.3	319	2	D71533 [acyl-carrier-prot
39	439.5	29.0	367	2	T00580 probable in polyke
40	426	28.1	650	2	C69678 probable polyketid
41	425	28.0	288	2	A69678 probable [acyl-ket
42	420.5	27.7	309	2	F71976 rifamycin polyketi
43	419	27.6	3413	2	T17467 probable [acyl-car
44	416.5	27.4	309	2	B64531 probable polyketid
45	414.5	27.3	1402	2	D70634

ALIGNMENTS

RESULT 1
F95048
malonyl CoA-acyl carrier protein transacylase [imported] - Streptococcus pneumoniae (st
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: F95048
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf
neon, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.C.; Venter, J.C.; Dougherty, B.A.; Morrisor
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95048
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-306 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK74583.1; PID:G14971890; GSPDB:GN00164; TIGR:SE
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0420
C:Superfamily: [acyl-carrier-protein] S-malonyltransferase; [acyl-carrier-protein] S-me

Query Match	100.0%	Score	1518	DB	2	Length	306
Best Local Similarity	100.0%	Pred. NO.	3.3e-99				
Matches	306	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	MTKTAFIFAGGQAQYLGMRDFYDQYPIVXETIDRASQVLGYDLRLYLIDTEEDKLNQTRY	60				
Db	1	MTKTAFIFAGGQAQYLGMRDFYDQYPIVXETIDRASQVLGYDLRLYLIDTEEDKLNQTRY	60				
Qy	61	TQPAILATSVAIRLLQEKGYQPDWAGLSLGEYSALVASGALDFEDAVALKRGAYME	120				
Db	61	TQPAILATSVAIRLLQEKGYQPDWAGLSLGEYSALVASGALDFEDAVALKRGAYME	120				
Qy	121	EAAPADSGKMWAVLNTPEVIEEACQKASELGVVTPANYNTPAQIVIAGEVWAVDRAVEL	180				
Db	121	EAAPADSGKMWAVLNTPEVIEEACQKASELGVVTPANYNTPAQIVIAGEVWAVDRAVEL	180				
Qy	181	LOEAGAKRLIPLKVGSPFFHTALLEPASQKLAETIAQVSPDSFTCPPLVGNTEAAVMQKEDI	240				
Db	181	LOEAGAKRLIPLKVGSPFFHTALLEPASQKLAETIAQVSPDSFTCPPLVGNTEAAVMQKEDI	240				
Qy	241	AQLLTROVKBPVRPYEYISIGVMQAGISNFIETGPGKVLGKLVGVKKIDQTAHLAHVEDQASL	300				
Db	241	AQLLTROVKBPVRPYEYISIGVMQAGISNFIETGPGKVLGKLVGVKKIDQTAHLAHVEDQASL	300				
Qy	301	VALLEK 306					
Db	301	VALLEK 306					

[illegible]

```

QY      299 SLVALLEK 306
      |
      ||:
Db      301 SPEALING 308

RESULT 4
G97338
A:Name: malonyl CoA-acyl carrier protein transacylase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision_14-Sep-2001 #text_change 30-Sep-2001
C:Accession: G97338
R:McNelling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum ATCC824
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: G97338
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-308 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK81498.1; PID:gl5026671; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC3575
C:Superfamily: [acyl-carrier-protein] S-malonyltransferase; [acyl-carrier-protein]
Query Match      47.0%   Score 713.5;  DB 2;  Length 308;

```

[illegible]

Nucleic Acids Res. 28, 4317-4331, 2000
A>Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
A/Reference number: A83650; MUID:20512582; PMID:11058132
A/Accession: D83961
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-313 <SPO>
A/Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BA806211.1; GSPDB:GN00
A/Experimental source: strain C-125
C/Genetics:
A/Gene: fabD
C/Superfamily: [acyl-carrier-protein] S-malonyltransferase; [acyl-carrier-protein] S-malonyltransferase
Query Match 43.1%; Score 654; DB 2; Length 313;
Best Local Similarity 45.6%; Pred. No. 1.3e-38;
Matches 141; Conservative 59; Mismatches 106; Indels 4; Gaps 4;
QY 1 MTKTAEFLPAGGQAGYLGMSDFDQYPIVKETIDRASQVGLYDL-RLYLDTEEDKLNQTR 59
DB 1 MAKVAFLEPGGQSGVMSGSELLSEK-AKEIFDEADERIGYLSLSIMFEGPEKLRTE 59
QY 60 YTPQAILATSVAVIRLLOEKGYQPMVAGLSLGEYSALVSGALDFEDAVLVAKGAYM 119
DB 60 NTPQALLTMSVAVIRLLOEKGYQPMVAGLSLGEYSALVSGALDFEDAVLVAKGAYM 119
QY 120 EEAAPADSGRWAVLNTPEVEIEACQKASELG-VVTPANYNTPAQIVIAGEVAVDRAV 178
DB 120 EEAAPADSGRWAVLNTPEVEIEACQKASELG-VVTPANYNTPAQIVIAGEVAVDRAV 178
QY 179 ELLOEAGAKRLIPLKVSPPHTALLESASQKLAETLAQVSDFTCPVLGNTAAVMQK- 237
DB 179 ELLOEAGAKRLIPLKVSPPHTALLESASQKLAETLAQVSDFTCPVLGNTAAVMQK- 237
QY 180 EEAKEAGAKRVIPLOVSGPPHSLMKPAEKLAVLADIAADAPPVIANVTADLVQKA 239
DB 180 EEAKEAGAKRVIPLOVSGPPHSLMKPAEKLAVLADIAADAPPVIANVTADLVQKA 239
QY 238 EDIAQLLTROVKEPVPFVYFISGVQWQAGISNFTPEIGSGKVLGFKKIDQTAHLAHVEDQ 297
DB 238 EDIAQLLTROVKEPVPFVYFISGVQWQAGISNFTPEIGSGKVLGFKKIDQTAHLAHVEDQ 297
QY 240 ADIRSSILIEGVSPVWEDTVRMLGLGDTFVEIGSGNLVSLVRKVRNVFVSVD 299
DB 240 ADIRSSILIEGVSPVWEDTVRMLGLGDTFVEIGSGNLVSLVRKVRNVFVSVD 299
QY 298 ASLVALLEK 306
DB 298 ASLVALLEK 306
QY 300 ASIEAMVKK 308
DB 300 ASIEAMVKK 308
RESULT 8
B41856
[acyl-carrier-protein] S-malonyltransferase (EC 2.3.1.39) fabD [validated] - Escherichia
C/Species: Escherichia coli
C/Date: 04-Mar-1993 #sequence revision 13-Jan-1995 #text_change 01-Mar-2002
C/Accession: B41856; A42147; S20443; A64853
R/Verwoert, J.I.; Verbrue, E.C.; van der Linden, K.H.; Nijtkamp, H.J.; Stuitje, A.R.
J. Bacteriol. 174, 2851-2857, 1992
A>Title: Cloning, nucleotide sequence, and expression of the *Escherichia coli* fabD gene,
A/Reference number: A41856; MUID:92234941; PMID:1314802
A/Accession: B41856
A/Molecule type: DNA
A/Residues: 1-309 <VER>
A/Cross-references: GB:M87040; NID:g145885; PIDN:AAA23742.1; PID:g145887
A/Note: Sequence extracted from NCBI backbone (NCBI:97135; NCBI:97148)
R/Rawlings, M.; Cronan Jr., J.B.
J. Biol. Chem. 267, 5751-5754, 1992
A>Title: The gene encoding *Escherichia coli* acyl carrier protein lies within a cluster
A/Reference number: A42147; MUID:92210530; PMID:1556094
A/Accession: A42147
A/Molecule type: DNA
A/Residues: 289-309 <RAW>
A/Cross-references: GB:M84891; NID:g145879; PIDN:AAA23738.1; PID:g145880
R/Magnuson, K.; Oh, W.; Larson, T.J.; Cronan Jr., J.B.
FEBS Lett. 299, 262-266, 1992
A>Title: Cloning and nucleotide sequence of the fabD gene encoding malonyl coenzyme A-
A/Reference number: S20443; MUID:92133950; PMID:1339356
A/Accession: S20443
A/Molecule type: DNA
A/Residues: 1-309 <WAG>
A/Cross-references: EMBL:211565; NID:g41363; PIDN:CAA77658.1; PID:g41364

A/Experimental source: strain K-12
A/Note: amino end of the mature protein confirmed by protein sequencing
R/Ruch, F.E.; Vagelos, P.R.
J. Biol. Chem. 248, 8095-8106, 1973
A>Title: Characterization of a malonyl-enzyme intermediate and identification of the ma
A/Reference number: A55383; MUID:74030718; PMID:4584823
A/Contents: annotation; active site
R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of *Escherichia coli* K-12.
A/Reference number: A64720; MUID:97426617; PMID:9278503
A/Accession: A64853
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-309 <BIAT>
A/Cross-references: GB:AE000210; GB:U00096; NID:g1787332; PIDN:AACT4176.1; PID:g1787334
A/Experimental source: strain K-12, substrain MG1655
C/Genetics:
A/Gene: fabD; tfpA
A/Map position: 24 min
C/function:
A/Description: catalyzes the reaction of malonyl-CoA with acyl carrier protein to form
A/Pathway: fatty acid biosynthesis
C/Superfamily: [acyl-carrier-protein] S-malonyltransferase; [acyl-carrier-protein] S-ma
C/Keywords: acyltransferase; coenzyme A; fatty acid biosynthesis
F:2-309/Product: [acyl-carrier-protein] S-malonyltransferase #status experimental <WAT>
F:5-289/Domain: [acyl-carrier-protein] S-malonyltransferase homology <WAT>
F:92/Active site: Ser (covalent substrate-binding) #status experimental
F:201/Active site: His #status predicted
Query Match 42.4%; Score 643.5; DB 1; Length 309;
Best Local Similarity 45.5%; Pred. No. 7.1e-38;
Matches 141; Conservative 60; Mismatches 102; Indels 7; Gaps 4;
QY 1 MTKTAEFLPAGGQAGYLGMSDFDQYPIVKETIDRASQVGLYDLRLYLDTEEDKLNQ 57
DB 1 MTKTAEFLPAGGQAGYLGMSDFDQYPIVKETIDRASQVGLYDLRLYLDTEEDKLNQ 57
QY 58 TRVTPAILATSVAVIRLLOEKGYQ-PMVAGLSLGEYSALVSGALDFEDAVLVAKRG 116
DB 58 TRVTPAILATSVAVIRLLOEKGYQ-PMVAGLSLGEYSALVSGALDFEDAVLVAKRG 116
QY 59 TWQTPALLTMSVAVIRLLOEKGYQ-PMVAGLSLGEYSALVSGALDFEDAVLVAKRG 118
DB 59 TWQTPALLTMSVAVIRLLOEKGYQ-PMVAGLSLGEYSALVSGALDFEDAVLVAKRG 118
QY 117 AYMEAPADSGRWAVLNTPEVEIEACQKASELG-VVTPANYNTPAQIVIAGEVAVDRAV 176
DB 117 AYMEAPADSGRWAVLNTPEVEIEACQKASELG-VVTPANYNTPAQIVIAGEVAVDRAV 176
QY 119 KFMQEAPEPTGMAALIGLDDASIAKACAEAGQVSPVNFSPQGVVIAHKEAVER 178
DB 119 KFMQEAPEPTGMAALIGLDDASIAKACAEAGQVSPVNFSPQGVVIAHKEAVER 178
QY 177 AVELLOEAGAKRLIPLKVSPPHTALLESASQKLAETLAQVSDFTCPVLGNTAAVMQ 236
DB 177 AVELLOEAGAKRLIPLKVSPPHTALLESASQKLAETLAQVSDFTCPVLGNTAAVMQ 236
QY 179 AGAACKAAGAKRVIPLOVSGPPHSLMKPAEKLAVLADIAADAPPVIANVTADLVQKA 238
DB 179 AGAACKAAGAKRVIPLOVSGPPHSLMKPAEKLAVLADIAADAPPVIANVTADLVQKA 238
QY 237 KED-IAQLLTROVKEPVPFVYFISGVQWQAGISNFTPEIGSGKVLGFKKIDQTAHLAHVE 295
DB 237 KED-IAQLLTROVKEPVPFVYFISGVQWQAGISNFTPEIGSGKVLGFKKIDQTAHLAHVE 295
QY 239 NGDAIRDALVRQLVNPVQWTKSVYMAAQGVHLYEVGPGKVLGTLTKRIVDTLTASALN 298
DB 239 NGDAIRDALVRQLVNPVQWTKSVYMAAQGVHLYEVGPGKVLGTLTKRIVDTLTASALN 298
QY 296 DOASLVALLE 305
DB 296 DOASLVALLE 305
QY 299 EPSAAMAALE 308
DB 299 EPSAAMAALE 308
RESULT 9
F90812
[acyl-carrier-protein] transacylase [imported] - Escherichia coli (strain C
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence revision 18-Jul-2001 #text_change 03-Aug-2001
C/Accession: F90812
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gen
A/Reference number: A99629; MUID:21156231; PMID:11258796
A/Accession: F90812
A/Status: preliminary

AI0194
[acyl-carrier-protein] S-malonyltransferase (EC 2.3.1.39) [imported] - Yersinia pestis
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 11-Jan-2002
C:Accession: AI0194
R:Farhail, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-rarrega, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AI0194
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-309 <XUR>
A:Cross-references: GB:AL590842; PIDN:CAC90420.1; PID:gl5979636; GSPDB:GN00175
C:Genetics:
A:Gene: fabD
C:Superfamily: [acyl-carrier-protein] S-malonyltransferase; [acyl-carrier-protein] S-mal
C:Keywords: acyltransferase; Coenzyme A

Query Match 41.4%; Score 628.5; DB 2; Length 309;
Best Local Similarity 45.5%; Pred. No. 8.1e-37;
Matches 140; Conservative 56; Mismatches 109; Indels 3; Gaps 3;
QY 1 MTKTAFLEAGGQAQYLGMRDFDYQYPIVKETIDRASQVLGYDLRYLIDT-EEDKLNQTR 59
Db 1 MSKFAVFPQGSQSLMLADLAQAQPIVEATFSEASSVLGYDLWLQVQGPABELNKTW 60
QY 60 YTOPAILATSVAIYRLLOEKGYO-PDMVAGLSGEYSALVAGSALDGFEDAVLAKRAY 118
Db 61 QTOPALLTASVAIYRLWQHGGKPLPSLMAGHSLGEYSALVAGSALDGFQKQAVSLVLRGL 120
QY 119 MEEAAPADSGKMWAVLNTPVEVIEEACQKASELGVTTPANYTTPAQIVIAAGEVAVDRAV 178
Db 121 MQEAVPEGTGAMTAIYGLNENSTAKACESEAGQVSPVNFNSPGQVVIAGNKDAVERAG 180
QY 179 ELQEAGAKELIPLKVSQGFHTALLESFASQKLAETLAQVSFSDTCTPLVGNTEA-ANMOK 237
Db 181 AACKAAGAKALPLVSPVSHCALMKPAADKLAVALDKIDFQEPPLFPVNVNVDKAEVSP 240
QY 238 EDIAQLLTQVKPEPVRFYFESIGVMQEAGISNFIETGPGKVLGSGFVKKIDQTAHLAHDQ 297
Db 241 VAIRSAVLQYLPVNFVRWTSVEFIAAEGVELLEIGPGKVLGLTKRIVDSLAARAVNDV 300
QY 298 ASLVALL 305
Db 301 VTLSALE 308

RESULT 13
AC0642
malonyl CoA-acyl carrier protein transacylase [imported] - Salmonella enterica subsp. en
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AC0642
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Conerton, P.; Cronin, A.; Davies, P.; Davies, R.N.; Dowd, L.; White, N.; Farrar,
S.; Mule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; PMID:11677608
A:Accession: AC0642
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-309 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08318.1; PID:gl6502364; GSPDB:GN00176
C:Genetics:
A:Gene: STY1233
C:Superfamily: [acyl-carrier-protein] S-malonyltransferase; [acyl-carrier-protein] S-mal

Query Match 40.3%; Score 612.5; DB 2; Length 309;
Best Local Similarity 43.9%; Pred. No. 1.1e-35;
Matches 139; Conservative 62; Mismatches 97; Indels 17; Gaps 5;
QY 1 MTKTAFLEAGGQAQYLGMRDFDYQYPIVKETIDRASQVLGYDLRYLIDTEE---DKLNQ 57
Db 1 MTQFAVFPQGSQSLMLAEMAANYPIVEETFAEASALGYDLMAL---TQGPABELANK 58
QY 58 TRYTOPAILATSVAIYRLLOEK-GYOPDMVAGLSGEYSALVAGSALDGFEDAVLAKRG 116
Db 59 TWOTQALLTASVALWRVWQGGKMPALMAGHSLGEYSALVAGSALVAGSALVAGSALVAGSALV 118
QY 117 AYEEAAPADSGKMWAVLNTPVEVIEEACQKASELGVTTPANYTTPAQIVIAAGEVAVDR 176
Db 119 KFQEAPEGTGMSALIGLDDASIAKACESEAGQVSPVNFNSPGQVVIAGHKEAVER 178
QY 177 AVELQEAGAKELIPLKVSQGFHTALLESFASQKLAETLAQVSFSDTCTPLVGN-----T 230
Db 179 AGAACKAAGAKALPLVSPVSHCALMKPTADKLAVALAKITFSAPTVPVNVNVDKVCET 238
QY 231 EAAVMQKEDIQALLTQVKPEPVRFYFESIGVMQEAGISNFIETGPGKVLGSGFVKKIDQTAH 290
Db 239 DAAA-----TRDALVQLYLPVNFVQWTKSVEFIAAAGVHELYEYVGPGLVGLTKRIVDTLT 293
QY 291 LAHVDDQASLVALL 304
Db 294 ASALNEPAALSAAL 307

RESULT 14
T44433
[acyl-carrier-protein] S-malonyltransferase (EC 2.3.1.39) fabD [similarity] - Moritella
C:Species: Moritella marina
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jun-2000
C:Accession: T44433
R:Morita, N.; Ueno, A.; Tanaka, M.; Ohgiya, S.; Hoshino, T.; Kawasaki, K.; Yumoto, I.;
Biotechnol. Lett. 21, 641-646, 1999
A:Title: Cloning and sequencing of clustered genes involved in fatty acid biosynthesis
A:Reference number: Z22768
A:Accession: T44433
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-307 <MOR>
A:Cross-references: EMBL:AB021978; PIDN:BAA85255.1
A:Experimental source: ATCC 15381
C:Genetics:
A:Gene: fabD
C:Superfamily: [acyl-carrier-protein] S-malonyltransferase; [acyl-carrier-protein] S-ma
C:Keywords: acyltransferase; coenzyme A
F:5-289/Domain; [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F:92/Active site: Ser (covalent substrate-binding) #status predicted
F:201/Active site: His #status predicted

Query Match 40.3%; Score 611.5; DB 2; Length 307;
Best Local Similarity 45.7%; Pred. No. 1.2e-35;
Matches 139; Conservative 59; Mismatches 101; Indels 5; Gaps 4;
QY 1 MTKTAFLEAGGQAQYLGMRDFDYQYPIVKETIDRASQVLGYDLRYLID--TEEDKLNQ 58
Db 1 MSKYAVFPQGSQSLMLADLAEEHVSVEQTPAQASEMLGYDLWLVDVHGHTVEE-LSQT 59
QY 59 RYTOPAILATSVAIYRLLOEK-GYOPDMVAGLSGEYSALVAGSALDGFEDAVLAKRG 117
Db 60 HITOPALLTASVALWRVIAAKEDFKPALVAGHSLGEYSALVAGSALVAGSALVAGSALV 119
QY 118 YNEEAAPADSGKMWAVLNTPVEVIEEACQKASELGVTTPANYTTPAQIVIAAGEVAVDRA 177
Db 120 LMQQAQVPOGICGAMAARVIGLDNDALIAACATAAEDEVVSAVNFNSPGQVVIAGNKAUVRA 179
QY 178 VELLQEAGAKELIPLKVSQGFHTALLESFASQKLAETLAQVSFSDTCTPLVGNTEA-AAVMQ 236
Db 180 SELCVAGARVMPPLVSPVSHCSLMKPADELKLALEKVTFTNTPVTKLINNVDAAPVD 239

Search completed: June 11, 2003, 20:04:36
Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 19:45:37 ; Search time 36 Seconds
(without alignments)
352.549 Million cell updates/sec

Title: US-09-308-397-2

Perfect score: 1518
Sequence: 1 MTKTFLPAGQAGYLONGR.....QTAHLAHDQASLVALLK 306

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	669.5	44.1	317	1 FABD_BACSU	P71019 bacillus su
2	638.5	42.1	308	1 FABD_ECOLI	P25715 escherichia
3	636	41.9	312	1 FABD_HAEIN	P25712 haemophilus
4	611.5	40.3	308	1 FABD_SALTY	O85140 salmonella
5	459	30.2	293	1 FABD_SYNY3	P73242 synechocyst
6	305.5	20.1	3491	1 ERY1_SACER	Q03131 saccharopol
7	274.5	18.1	302	1 FABD_MYCTU	Q10501 mycobacteri
8	269	17.7	2181	1 STCA_EMENI	Q12397 emericella
9	266.5	17.6	1986	1 WGENI	Q03149 emericella
10	254	16.7	2109	1 PXS1_ASPPA	Q12053 aspergillus
11	253.5	16.7	3519	1 OLS6_STRAT	Q07017 streptomyce
12	246	16.2	3567	1 ERY2_SACER	Q03132 saccharopol
13	242.5	16.0	3172	1 ERY3_SACER	Q03133 saccharopol
14	224	14.8	1538	1 PSB_MYCTU	Q10978 mycobacteri
15	208.5	13.7	1876	1 PSB_MYCTU	Q10977 mycobacteri
16	205	13.5	2037	1 FAS1_CANAL	P34711 c fatty aci
17	199.5	13.1	2505	1 FAS1_FAT	P12785 rattus norv
18	196.5	12.9	2076	1 FAS1_YARLI	P34229 y fatty aci
19	195.5	12.9	1774	1 MGAS_PENPA	P22367 penicillium
20	186	12.3	2110	1 MCAS_MYCBO	Q02251 mycobacteri
21	184	12.1	2051	1 FAS1_YEAST	P07149 s fatty aci
22	173.5	11.4	2073	1 FAS1_SCHPO	Q8uug0 s fatty aci
23	169	11.1	2504	1 FAS1_HUMAN	P49327 homo sapien
24	167.5	11.0	2504	1 FAS1_CHICK	P12276 gallus gall
25	149.5	9.8	1914	1 STCK_EMENI	Q00706 emericella
26	108	7.1	398	1 PKK_LACLA	Q00706 emericella
27	101.5	6.7	563	1 ENLI_ORYGA	P17814 oryza sativ
28	99.5	6.6	431	1 CLO1_CLOAB	Q97152 clostridium
29	99.5	6.6	502	1 ATPA_GUITH	O78475 guillardia
30	99.5	6.6	505	1 ATPA_CYAPA	P28080 cyanophora
31	99.5	6.6	506	1 ATPA_ANASP	P2405 cyanobact sp
32	98.5	6.5	756	1 YASI_RHISN	P5652 rhizobium s
33	97.5	6.4	3396	1 POLG_DENIS	P33478 d genome po

RESULT 1

ID	FABD_BACSU	STANDARD	PRT	317 AA
AC	P71019: O34463			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39) (MCT).			
GN	FABD.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=98195738; PubMed=9534248;			
RA	Foulger D., Errington J.,			
RT	"A 28 kbp segment from the spoVM region of the Bacillus subtilis 168			
RT	Genome."			
RL	Microbiology 144:801-805(1998).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=98044033; PubMed=9384377;			
RA	Kunst F., Ogatawara N., Moser I., Albertini A.M., Alloni G.,			
RA	Azevedo V., Bertorello M.G., Bessieres P., Bolotin A., Borchert S.,			
RA	Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,			
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,			
RA	Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,			
RA	Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,			
RA	Entian K.D., Errington J., Fabelt C., Ferrari E., Foulger D.,			
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,			
RA	Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,			
RA	Guiseppi G., Guy B.J., Haga K., Haleth J., Harwood C.R., Henaut A.,			
RA	Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,			
RA	Joris B., Karamata D., Kasahara Y., Klaer-Bianchard M., Klein C.,			
RA	Kobayashi Y., Koetter P., Koningstein G., Krcgh S., Kumano M.,			
RA	Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,			
RA	Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,			
RA	Medina N., Mellado R.P., Mizuno M., Noestl D., Nakai S., Noback M.,			
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,			
RA	Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,			
RA	Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,			
RA	Rieger M., Rivolta C., Roche E., Roche M., Rose M., Sadale Y.,			
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,			
RA	Sekiuchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,			
RA	Sorokin A., Taconetti A., Takagi T., Takahashi H., Takemaru K.,			
RA	Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,			

ALIGNMENTS

34	97	6.4	872	1	SYA_LACLA
35	97	6.4	1195	1	METH_SYNY3
36	97	6.4	1305	1	RFOC_UREPA
37	96.5	6.4	876	1	DFO1_BACST
38	95.5	6.3	446	1	EX7L_STRPN
39	95	6.3	404	1	HISX_ARCFU
40	94.5	6.2	357	1	BUK_THETN
41	94.5	6.2	442	1	MCRB_METTH
42	94.5	6.2	604	1	GLMS_LACLA
43	94.5	6.2	1148	1	YKQ0_CABEL
44	94.5	6.2	1357	1	RPOB_PSEPU
45	94	6.2	323	1	YD95_ARCFU

Q9cew0	lactococcus
Q95786	synechocyst
Q9pqv5	ureaplasma
P52026	bacillus st
Q97qj8	streptococc
Q30027	archaeoglob
Q8r832	thermoanaer
O27236	methanobact
O9cgt5	l glucosami
P34305	caenothabdi
P91975	pseudomonas
O28876	archaeoglob

RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumsstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*,"
RL Nature 390:249-256(1997).
CC -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC malonyl-[acyl-carrier protein].
CC -!- PATHWAY: Fatty acid biosynthesis.
CC -!- SIMILARITY: SIGNIFICANT, TO SEVERAL OTHER PROTEINS WITH
CC TRANSACYLASE ACTIVITY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U59433; AAC44306.1; -;
CC EMBL; Y13937; CAA74249.1; -;
CC EMBL; Z99112; CAB13463.1; -;
CC HSSP; P25715; 1MLA.
CC Subtilist; BG11836; fabD.
CC InterPro; IPR001227; Ac transferase.
CC InterPro; IPR004410; fabD.
CC Pfam; PF00698; Acyl_transf; 1.
CC TIGRFAMs; TIGR00128; fabD; 1.
CC Fatty acid biosynthesis; Transferase; Complete proteome.
CC BY SIMILARITY.
CC ACT_SITE 91 91
CC FT ACT_SITE 201 201 VPAGEGAM -> GCRLAKEOM (IN REF. 1).
CC FT CONFLICT 124 131
CC SQ SEQUENCE 317 AA; 34035 MW; 191AE82881C91D7F CRC64;

Query Match 44.1%; Score 669.5; DB 1; Length 317;
Best Local Similarity 47.2%; Pred. No. 3.8e-39;
Matches 143; Conservative 47; Mismatches 110; Indels 3; Gaps 3;

QY 1 MKTAFIFAGGAGVGLMGDFDYDQYPIVKETIDRASVLGYDLRYLI-DTEEDKLNQTR 59
Db 1 MSKTAFLFPGGSGFINGWELHEQVPAARLFDDEADLETLSLIFEGDAEETLTLY 60
QY 60 YTPAILATSVAIYRLLOEKGYQDMVAGLSLGEYSALVAGSLDFEDAVLVAKRGAYM 119
Db 61 NAOQALLTTSIAVLEKFKESGITPDFTAGHSLGEYSALVAAGLSFKDAVYTVRKRGEPF 120
QY 120 EERAPADSGKMAVLNTPFVEIEACOKASELG-VYTPPNYNTPAQIVTAGEVVAVDRAV 178
Db 121 NEAVPAGEGAAALIGMDAALKQVTDKVTTEGNVLQLANLPCQQLVISGTAKGVELAS 180
QY 179 ELLOEAGAKRLIPLKVSQPHFTALLEPASOKLAETLAQVFSDFCTPLVGNTEAAVM-QK 237
Db 181 ELAKENGAKRAIPLVSPSPHSELMFAEKLEVLDAACDIKDADVPLSNVSADVWTEK 240
QY 238 EDIAQLITRQVKEVRYEYIGVQWQAGISNFIIEGKVLGFKVKKIDOTAHAVEQ 297
Db 241 ADIKEKLEIQLYSPVFEESINKLIAEGVTFTEIGPKVLGLVKVNRRLKTIIVASDP 300
QY 298 ASL 300
Db 301 ETI 303

RESULT 2
FABD_ECOLI STANDARD; PRT; 308 AA.
ID FABD_ECOLI
AC P25715;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39) (MCT).

GN FABD OR TPPA OR B1092.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC STRAIN=K12;
RX MEDLINE=92183950; PubMed=1339356;
RA Magnuson K., Oh W., Larson T.J., Cronan J.E. Jr.;
RT "Cloning and nucleotide sequence of the fabD gene encoding malonyl
RL coenzyme A-acyl carrier protein transacylase of Escherichia coli,";
RN FEBS Lett. 299:262-266(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92234941; PubMed=1314802;
RA Verwoert I.I., Verbree E.C., van der Linden K.H., Nijkamp H.J.,
RT Stuitje A.R.;
RN "Cloning, nucleotide sequence, and expression of the Escherichia coli
RT fabD gene encoding malonyl coenzyme A-acyl carrier protein
RT transacylase,";
RL J. Bacteriol. 174:2851-2857(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95214624; PubMed=7700236;
RA Bouquin N., Tempete M., Holland I.B., Seror S.J.;
RT "Resistance to trifluoroperazine, a calmodulin inhibitor, maps to the
RT fabD locus in Escherichia coli,";
RL Mol. Gen. Genet. 246:628-637(1995).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden W.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12,";
RL Science 277:1453-1474(1997).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map,";
RL DNA Res. 3:137-155(1996).
RN [6]
RP SEQUENCE OF 288-308 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92210530; PubMed=1556094;
RA Cronan J.E. Jr., Rawlings M.;
RT "The gene encoding Escherichia coli acyl carrier protein lies within
RT a cluster of fatty acid biosynthetic genes,";
RL J. Biol. Chem. 267:5751-5754(1992).
RN [7]
RP SEQUENCE OF 1-11.
RC STRAIN=K12 / W3110;
RA Pasquali C., Sanchez J.-C., Ravier F., Golaz O., Hughes G.J.,
RA Frutiger S., Paquet N., Wilkins M., Appel R.D., Bairoch A.,
RA Hochstrasser D.F.;
RL Submitted (SEP-1994) to the SWISS-PROT data bank.
RN [8]
RP SEQUENCE OF 1-10.
RC STRAIN=K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded

in the genome of *Escherichia coli* K-12.";
Electrophoresis 18:1259-1313 (1997).
[9]
RN
X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RX
MEDLINE=95286570; PubMed=7768883;
RR
Serre L., Verbrée E.C., Dauter Z., Stuitje A.R., Derewenda Z.S.;
RT "The *Escherichia coli* malonyl-CoA:acyl carrier protein transacylase
RT at 1.5-A resolution. Crystal structure of a fatty acid synthase
RT component";
RT J. Biol. Chem. 270:12961-12964 (1995).
RL
CC -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC malonyl-[acyl-carrier protein].
CC -!- PATHWAY: Fatty acid biosynthesis.
CC -!- SIMILARITY: SIGNIFICANT, TO SEVERAL OTHER PROTEINS WITH
CC TRANSACYLASE ACTIVITY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M87040; AAC23742.1; -
CC EMBL; Z11565; CAAT7658.1; -
CC EMBL; M84991; AAA23738.1; -
CC EMBL; A5000210; AAC74176.1; -
CC EMBL; D30745; BAA35900.1; -
CC PIR; S20443; S20443.
CC PIR; B41856; B41856.
CC PDB; 1MLA; 25-JAN-95.
CC SWISS-2DPAGE; P25715; COLI.
CC EcoGene; E811317; fabD.
CC InterPro; IPR001227; AC_transferase.
CC InterPro; IPR004410; FabD.
CC Pfam; PF00698; Acyl_transf_1.
CC TIGRFAMs; TIGR00128; fabd; 1.
CC Fatty acid biosynthesis; Transferase; 3D-structure; Complete proteome.
CC INIT_MET 0 0
CC FT ACT_SITE 91 91
CC FT ACT_SITE 200 200
CC FT ACT_SITE 308 AA; 32286 MW; F5901043D92FED8E CRC64;
CC SEQUENCE 308 AA; 32286 MW; 5901043D92FED8E CRC64;
Query Match 42.1%; Score 638.5; DB 1; Length 308;
Best Local Similarity 45.3%; Pred. No. 5e-37;
Matches 140; Conservative 60; Mismatches 102; Indels 7; Gaps 4;
QY 2 TKTAFLPAGCAQYLGMGRDFDYDPIVKETIDRASQVGLDLYRLIDTEE---DKLNOT 58
DB 1 TQFAFLPFGQGSQTQVGMADMAASPIVEETFAESALGVDLMAL--TQGGFAEELNKT 58
QY 59 RYTPQAILATSAIVAYRLLOEKGYQ--PDMVAGLSLGEYSALVASGALDFEDAVALKRGA 117
DB 59 WQTPALLTASVALYRVWQQGKAPAMWAGHSLGEYSALVCAGVIDFADAVRLVEMRGK 118
QY 118 YMEEAAPADSGKRWAVLNTFVEIEEACQKASELGVWTPANVTPTAQIVIAEVAVDRA 177
DB 119 FMOEAIVPEGTGVAALIGLDLDSIAKACBAAEGGVSPVNFSPGVVIAHKEAVERA 178
QY 178 VELLQEAGAKRLIPLKVSQPFHTALLEPASQKLAETLAQVSFSDPTCLPVGNTEAAVMQK 237
DB 179 GAACKAAGAKRALPLFVSVPFHSCHALMKPAADKLAVELAKITFNATFVFWNNVDVKCTN 238
QY 238 ED-IAGLLTRQVKEPVRFFVSTGVQMOEAGISNFIIGPKVLSGFVKIDQTAHLAVED 296
DB 239 GDAIRDALVRQLXNPQVTKSVEYMAAQGVHELYEVGPGKVLTLGTLTKRIVDTLTASALNE 298
QY 297 QASLVALLE 305
DB 299 PSAMAAALE 307

Db 181 AALCKEAGAKRALPLAVSFHSCALMKPAAEQALAVLENIQINTPTISVNNVDVKAETE 240
 Qy 237 KEDIAQLLTQVKEPVRFYFESIGVMQAGISNFIETGPGKVLGSGFKKIDQTAHLAHVED 296
 Db 241 GTEIRTAIVQLVSPVETWEKMAQDGVLVLAEVGPGKVLNGLTKRIVGDLQAISVND 300
 Qy 297 QASLVALLE 305
 Db 301 VASFNAVEE 309

RESULT 4

FABD_SALTY
 ID FABD_SALTY STANDARD; PRT; 308 AA.
 AC 085140;
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39) (MCT).
 GN FABD OR STM1194;
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 CX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2;
 RX MEDLINE=98317265; PubMed=9642179;
 RA Zhang Y., Cronan J.E. Jr.;
 RT "Transcriptional analysis of essential genes of the Escherichia coli
 fatty acid biosynthesis gene cluster by functional replacement with
 the analogous Salmonella typhimurium gene cluster.";
 RJ J. Bacteriol. 180:3295-3303(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McLelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2.";
 RL Nature 413:852-856(2001).
 CC -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
 malonyl-[acyl-carrier protein].
 CC -!- PATHWAY: Fatty acid biosynthesis.
 CC -!- SIMILARITY: SIGNIFICANT, TO SEVERAL OTHER PROTEINS WITH
 TRANSACYLASE ACTIVITY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; AF044668; AAC38649.1; -;
 CC EMBL; AE008752; AAL20123.1; -;
 CC HSSP; P25715; 1MLA;
 CC StyGene; SG1637; fabD.
 DR InterPro; IPR001227; Ac transferase.
 DR InterPro; IPR004410; FabD.
 DR Pfam; PF00698; Acyl transf. 1.
 DR TIGRFAMs; TIGR00128; fabD; 1.
 KW Fatty acid biosynthesis; Transferase; Complete proteome.
 FT INIT MET 0
 FT ACT_SITE 91 91 BY SIMILARITY.
 FT ACT_SITE 200 200 BY SIMILARITY.
 SQ SEQUENCE 308 AA; 32274 MW; B3565B45341A611A CRC64;

Query Match 40.3%; Score 611.5; DB 1; Length 308;
 Best Local Similarity 44.1%; Pred. No. 3.5e-35;
 Matches 138; Conservative 62; Mismatches 96; Indels 17; Gaps 5;
 QY 2 TKTAFLGAGQAQVLGMRDFFDYDPIVKETIDRASQVLGYDLRLYLIDTEE---DKLNQT 58
 DB 1 TQFAFVFGGQSGVGLAEMAANYPIVETFAESAALGYDLWAL---TQGGPAEELNKT 58
 QY 59 RYTOPAILATSVAIYRLLOEK-GYQPDWVAGLSLGEYSALVASGALDFEDAVVALVAKGA 117
 DB 59 WQTOPALLTASVALRWYMQQGGKVPALMAGHSLGEYSALVCAGVINFAFVRLVVRGK 118
 QY 118 YMEEAAPADSGKMAVAVLNTPVEVIEEACOKASELGVVTTPANYNTPAQIVTAGEVVAVDRA 177
 DB 119 FMQEA VPEGTGMGSAIIGLDDASIAKACBSAEQGVVSPVNFSPGQVVIAGHKEAVERA 178
 QY 178 VELLQEAQAKELIFPKVSGPHTALLPASOKLAETLAQVSFSDFTCLPVGN-----TE 231
 DB 179 GAACKAAGAKRALPLVPVSPHSCALMKPAAADKLAELAKITFSAPTVPVNNVDVKCETD 238
 QY 232 AAVQKEDIQOLLTROVKEPVRFYFESIGVMQAGISNFIETGPGKVLGSGFKKIDQTAHL 291
 DB 239 AAA-----IRDALVRQLYNPVQWTKSVEFTAAQGVHELYEVGPGKVLTKRIVDTLTA 293
 QY 292 AHVEDQASLVALL 304
 DB 294 SALNEPAALSAL 306

RESULT 5

FABD_SYN3
 ID FABD_SYN3 STANDARD; PRT; 293 AA.
 AC P73242;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39) (MCT).
 GN FABD OR SUR2023.
 OS Synecocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
 CX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsumoto A., Muraki A., Nakazaki N., Naruo K.,
 RA Okumura S., Shimoda S., Takeuchi C., Wada T., Watanabe A.,
 RA Yamada M., Yasuda M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 Synecocystis sp. strain PCC6803. II. Sequence determination of the
 entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
 malonyl-[acyl-carrier protein].
 CC -!- PATHWAY: Fatty acid biosynthesis.
 CC -!- SIMILARITY: SIGNIFICANT, TO SEVERAL OTHER PROTEINS WITH
 TRANSACYLASE ACTIVITY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; D90904; BAA17269.1; -;
 CC HSSP; P25715; 1MLA;
 CC InterPro; IPR001227; Ac transferase.
 CC InterPro; IPR004410; FabD.
 CC Pfam; PF00698; Acyl transf. 1.
 CC TIGRFAMs; TIGR00128; fabD; 1.

KW Paddy acid biosynthesis; Transferase; Complete proteome.
 FT ACT SITE 88 BY SIMILARITY.
 FT ACT SITE 188 BY SIMILARITY.
 SQ SEQUENCE 293 AA; 31491 MW; 31B4ABA59ECDB377 CRC64;
 Query Match 30.2%; Score 459; DB 1; Length 293;
 Best Local Similarity 36.8%; Pred No. 9.8e-25;
 Matches 111; Conservative 59; Mismatches 118; Indels 14; Gaps 6;
 QY 3 KTAFLFAGCAQYQLGMRDFYDQVPIVKTETIDRASQVLYGLRYLITDEPKLNQRYTQ 62
 Db 2 KTAFLFAGCAQYQLGMRDFYDQVPIVKTETIDRASQVLYGLRYLITDEPKLNQRYTQ 60
 QY 63 PAILLATSVAILLQKQVQDMVAGLSGEYSALVAGSLDGEDAVALVAKRGAYVEEA 122
 Db 61 PCLYVIEAILADLLKQGFQDDYVAGHSLGEYSALVAGSLDGEDAVALVAKRGAYVEEA 117
 QY 123 APADSGKMAVLNTEVEIEACOKASELGVVVTNPANTPAQIVIAGEVAVDRAVELLQ 182
 Db 118 ASASGMMALMKPQDTLQQAALTNTDTE--VVLANDNSPQVVISGTVAGVE--AALA 171
 QY 183 EGAKRLPLKVGSPFHTALLEPASOKLAETLAQVSDFTCPVLGNTEAAVMQKED-IA 241
 Db 172 NVKARVPLKVGSPFHTALLEPASOKLAETLAQVSDFTCPVLGNTEAAVMQKED-IA 231
 QY 242 QLLTQVKEPVRVFSISGMQDEAGISNFIETIGPGKVLGSGFVKIDQTAHLAHV---EDQA 298
 Db 232 EKLIQMTGVSVRWRTWNLGEIGATDYWEVPGKVLGSGFVKIDQTAHLAHV---EDQA 291
 QY 299 SL 300
 Db 292 SL 293
 RESULT 6
 ERYL_SACER STANDARD; PRT; 3491 AA.
 ID ERYL_SACER
 AC Q03131;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Erythronolide synthase, modules 1 and 2 (EC 2.3.1.94) (ORF 1) (6-
 DE deoxyerythronolide B synthase I) (DESS 1).
 GN ERYA.
 OS Saccharopolyspora erythraea (Streptomyces erythraeus).
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Pseudonocardiaceae; Pseudonocardiaceae;
 OC Saccharopolyspora.
 OX NCBI_TaxID=1836;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91220065; PubMed=2024119;
 RA Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
 RT "Modular organization of genes required for complex polyketide
 biosynthesis.";
 RL Science 252:675-679 (1991).
 RN [2]
 RP SEQUENCE OF 3474-3491 FROM N.A.
 RX MEDLINE=93231529; PubMed=8386127;
 RA Donadio S., Staver M.J.;
 RT "IS1136, an insertion element in the erythromycin gene cluster of
 Saccharopolyspora erythraea.";
 RL Gene 126:147-151 (1993).
 CC -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
 CC deoxyerythronolide B.
 CC -!- COFACTOR: NADP; CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTHETHEINES.
 CC -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
 CC BIOSYNTHESIS.
 CC -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
 CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
 CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
 CC SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
 CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
 AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
 RESPECTIVELY.
 CC -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
 CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
 CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
 CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
 CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
 CC OF THE FULL-LENGTH CHAIN.
 CC -!- SIMILARITY: TO PATTY ACID SYNTHASE (PAS).
 CC -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC EMBL; MG3676; AAA26493.2; -;
 DR EMBL; L07626; AAA26504.1; -;
 DR HSSP; P25715; 1MLA.
 DR InterPro; IPR002198; ADH short.
 DR InterPro; IPR001227; Ac_Transferase.
 DR InterPro; IPR000794; Ketoacyl-synt.
 DR InterPro; IPR003880; Ppantne attach.
 DR Pfam; PF00106; adh short; 1.
 DR Pfam; PF00109; ketoacyl-synt; 2.
 DR Pfam; PF00550; pp-binding; 3.
 DR Pfam; PF00698; Acyl_transf; 3.
 DR Pfam; PF02801; ketoacyl-synt; 2.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 3.
 DR PROSITE; PS00606; B_KETOACYL SYNTHASE; 2.
 DR PROSITE; PS0075; ACP DOMAIN; 3.
 KW Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
 KW Phosphopantetheine; Multifunctional enzyme.
 FT DOMAIN 1 1972 MODULE 1.
 FT DOMAIN 1979 3491 MODULE 2.
 FT DOMAIN 1 375 ACYLTRANSFERASE (AT) 1.
 FT DOMAIN 414 484 ACYL CARRIER (ACP) 1.
 FT DOMAIN 503 961 BETA-KETOACYL SYNTHASE 1.
 FT DOMAIN 1030 1356 ACYLTRANSFERASE (AT) 2.
 FT DOMAIN 1611 1794 BETA-KETOACYL REDUCTASE 1.
 FT DOMAIN 1888 1958 ACYL CARRIER (ACP) 2.
 FT DOMAIN 1979 2441 BETA-KETOACYL SYNTHASE 2.
 FT DOMAIN 2507 2854 ACYLTRANSFERASE (AT) 3.
 FT DOMAIN 3055 3237 BETA-KETOACYL REDUCTASE 2.
 FT DOMAIN 3334 3404 ACYL CARRIER (ACP) 3.
 FT ACT SITE 145 145 ACYL-ENZYME INTERMEDIATE.
 FT BINDING 447 447 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT ACT_SITE 677 677 BETA-KETOACYL SYNTHASE.
 FT ACT_SITE 1128 1128 ACYL-ENZYME INTERMEDIATE.
 FT NP_BIND 1614 1660 NADP.
 FT BINDING 1921 1921 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT ACT_SITE 2148 2148 BETA-KETOACYL SYNTHASE.
 FT ACT_SITE 2598 2598 ACYL-ENZYME INTERMEDIATE.
 FT NP_BIND 3058 3104 NADP.
 FT BINDING 3367 3367 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 SQ SEQUENCE 3491 AA; 365022 MW; 652BFC32C90FA8C4 CRC64;
 Query Match 20.1%; Score 305.5; DB 1; Length 3491;
 Best Local Similarity 29.2%; Pred. No. 7.7e-13;
 Matches 93; Conservative 58; Mismatches 134; Indels 33; Gaps 10;
 QY 4 TATLFAQCAQYQLGMRDFYDQVPIVKTETIDRASQVLYGLRYLITDEPKLNQRYTQ 57
 Db 2507 SVFVFFGCAQCAQYQLGMRDFYDQVPIVKTETIDRASQVLYGLRYLITDEPKLNQRYTQ 56
 QY 58 TRYQPAILLATSVAILLQKQVQDMVAGLSGEYSALVAGSLDGEDAVALVAKRGAYVEEA 117
 Db 2566 VDVVQVFLFVAVVYSLAEALWACGAVPSAVIGHSGEIAAVALVAGSLDGEDAVALVAKRGAYVEEA 117
 QY 118 YMEERAPADSGKMAVLNTEVEIEACOKASELGVVVTNPANTPAQIVIAGEVAVDRAVELLQ 177

```

DR Tuberculin; RV2243; -.
DR InterPro: IPR001227; Acyl transferase.
DR Pfam: PF00698; Acyl transf, 1.
KW Fatty acid biosynthesis; Transferase; Complete proteome.
FT ACT_SITE 91 91 BY SIMILARITY.
FT ACT_SITE 194 194 BY SIMILARITY.
FT DOMAIN 206 210 POLY-ALA.
FT DOMAIN 233 236 POLY-ALA.
SQ SEQUENCE 302 AA; 30788 MW; BB7BCD8217FC66C8 CRC64;

Query Match 18.1%; Score 274.5; DB 1; Length 302;
Best Local Similarity 29.9%; Pred. No. 4.8e-12;
Matches 92; Conservative 52; Mismatches 135; Indels 29; Gaps 9;

QY 5 AFLFAGGAGYQLGMRDFYDQYPIVKTETIDRAQGVLYDLYLIDT-EEDKLNQTRYTOP 63
DB 3 ALLAPGQSGTQTEGMLSPWL-QLFCAADQIAAWKAADLDLARLGTASTEEITDTVAQP 61
QY 64 AILATSVAIYRLLEK- -GYQPMVAGLSLGEYSALVASGALDFEDAVALKRGAYME 120
DB 62 LIVATLILAHQELARRCVLAGKQVIVAGHSVGLIAYAYAGVIAADDAVALAATRGEMA 121
QY 121 EAAPADSGKMAVINTVEVIEBACQKASLGVTVPANNYTPAQIVIAGEVAVDRAVEL 180
DB 122 KACATEPTGMSAVLGSD- -TEVLSRLQDLV-PANRNAGQIVAAAGRLTALEK- -L 174
QY 181 LQERAGK-RLIPLVKSGPHTLLEPASOKLAETLAQVSPDFTCPVGNTEAAVMQKED 239
DB 175 AEDPPAKRVALGVAGAFHTEFMAPALDGFAAANIANIATADTATLLSNRDKPVTSA 234
QY 240 IA-QLLTRQVKEPVRFYESIGVMQEAGISNFIIGPKVLSGFVKX- - - - - 284
DB 235 AAMDVLVSQLTQPVWMDLCTATLREHTVTAIVEFPFAGTSLGIAKRELGVPARAVKSPA 294
QY 285 -IDQTAHL 291
DB 295 DLDELALN 302

RESULT 8
STCA_EMENI STANDARD; PRT; 2181 AA.
ID STCA_EMENI
AC Q12397;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative sterigmatocystin biosynthesis polyketide synthase (PKS).
GN STCA OR PKSST.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 26;
RX MEDLINE=98202293; PubMed=8643646;
RA Brown D.W., Yu J.-H., Kellar H.S., Fernandes M., Nesbitt T.C.,
RA Kellar N.P., Adams T.H., Leonard T.J.;
RA "Twenty-five coregulated transcripts define a sterigmatocystin gene
RT cluster in Aspergillus nidulans."
RL Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 4;
RX MEDLINE=95370159; PubMed=7642507;
RA Yu J.-H., Leonard T.J.;
RA "Sterigmatocystin biosynthesis in Aspergillus nidulans requires a
RT novel type I polyketide synthase."
RL J. Bacteriol. 177:4792-4800 (1995).
CC -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF
CC STERIGMATOCYSTIN FROM HEXANOYL-COA AND SEVEN MALONATES
CC -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTHETINES
CC (POTENTIAL).

```

```

DR Tuberculin; RV2243; -.
DR InterPro: IPR001227; Acyl transferase.
DR Pfam: PF00698; Acyl transf, 1.
KW Fatty acid biosynthesis; Transferase; Complete proteome.
FT ACT_SITE 91 91 BY SIMILARITY.
FT ACT_SITE 194 194 BY SIMILARITY.
FT DOMAIN 206 210 POLY-ALA.
FT DOMAIN 233 236 POLY-ALA.
SQ SEQUENCE 302 AA; 30788 MW; BB7BCD8217FC66C8 CRC64;

Query Match 18.1%; Score 274.5; DB 1; Length 302;
Best Local Similarity 29.9%; Pred. No. 4.8e-12;
Matches 92; Conservative 52; Mismatches 135; Indels 29; Gaps 9;

QY 5 AFLFAGGAGYQLGMRDFYDQYPIVKTETIDRAQGVLYDLYLIDT-EEDKLNQTRYTOP 63
DB 3 ALLAPGQSGTQTEGMLSPWL-QLFCAADQIAAWKAADLDLARLGTASTEEITDTVAQP 61
QY 64 AILATSVAIYRLLEK- -GYQPMVAGLSLGEYSALVASGALDFEDAVALKRGAYME 120
DB 62 LIVATLILAHQELARRCVLAGKQVIVAGHSVGLIAYAYAGVIAADDAVALAATRGEMA 121
QY 121 EAAPADSGKMAVINTVEVIEBACQKASLGVTVPANNYTPAQIVIAGEVAVDRAVEL 180
DB 122 KACATEPTGMSAVLGSD- -TEVLSRLQDLV-PANRNAGQIVAAAGRLTALEK- -L 174
QY 181 LQERAGK-RLIPLVKSGPHTLLEPASOKLAETLAQVSPDFTCPVGNTEAAVMQKED 239
DB 175 AEDPPAKRVALGVAGAFHTEFMAPALDGFAAANIANIATADTATLLSNRDKPVTSA 234
QY 240 IA-QLLTRQVKEPVRFYESIGVMQEAGISNFIIGPKVLSGFVKX- - - - - 284
DB 235 AAMDVLVSQLTQPVWMDLCTATLREHTVTAIVEFPFAGTSLGIAKRELGVPARAVKSPA 294
QY 285 -IDQTAHL 291
DB 295 DLDELALN 302

RESULT 8
STCA_EMENI STANDARD; PRT; 2181 AA.
ID STCA_EMENI
AC Q12397;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative sterigmatocystin biosynthesis polyketide synthase (PKS).
GN STCA OR PKSST.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 26;
RX MEDLINE=98202293; PubMed=8643646;
RA Brown D.W., Yu J.-H., Kellar H.S., Fernandes M., Nesbitt T.C.,
RA Kellar N.P., Adams T.H., Leonard T.J.;
RA "Twenty-five coregulated transcripts define a sterigmatocystin gene
RT cluster in Aspergillus nidulans."
RL Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 4;
RX MEDLINE=95370159; PubMed=7642507;
RA Yu J.-H., Leonard T.J.;
RA "Sterigmatocystin biosynthesis in Aspergillus nidulans requires a
RT novel type I polyketide synthase."
RL J. Bacteriol. 177:4792-4800 (1995).
CC -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF
CC STERIGMATOCYSTIN FROM HEXANOYL-COA AND SEVEN MALONATES
CC -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTHETINES
CC (POTENTIAL).

```

CC -1- PATHWAY: Sterigmatocystin biosynthesis; first step.
CC -1- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U34740; AAC49191.1; -;
CC EMEL; L39121; AAA91586.1; -;
CC HSP; P25715; LMLA.
CC InterPro: IPR001227; AC transferase.
CC InterPro: IPR000794; Ketoacyl-synt.
CC InterPro: IPR003880; Peptide attach.
CC InterPro: IPR001031; Thioesterase.
CC Pfam; PF00109; ketoacyl-synt; 1.
CC Pfam; PF00698; Acyl trans; 1.
CC Pfam; PF00975; Thioesterase; 1.
CC Pfam; PF02801; ketoacyl-synt C; 1.
CC PROSITE; PS50075; ACP DOMAIN; 2.
CC TRANSFERASE; Acyltransferase; Phosphopantetheine; Repeat;
CC Multifunctional enzyme.
CC FT DOMAIN 383 814 BETA-KETOACYL SYNTHASE.
CC FT DOMAIN 884 1209 ACYL/MALONYL TRANSFERASES.
CC FT DOMAIN 1706 1777 ACYL CARRIER (ACP) 1.
CC FT DOMAIN 1830 1901 ACYL CARRIER (ACP) 2.
CC FT DOMAIN ? 2181 THIOESTERASE.
CC FT ACT_SITE 552 552 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
CC FT ACT_SITE 978 978 ACYL/MALONYL TRANSFERASES (BY
CC SIMILARITY).
CC FT BINDING 1738 1738 PHOSPHOPANTHEINE (BY SIMILARITY).
CC FT BINDING 1862 1862 PHOSPHOPANTHEINE (BY SIMILARITY).
CC FT ACT_SITE 2028 2028 THIOESTERASE (BY SIMILARITY).
CC SQ SEQUENCE 2181 AA; 238831 MW; 5A3E5712AA9AD942 CRC64;
Query Match 17.7%; Score 269; DB 1; Length 2181;
Best Local Similarity 27.0%; Pred. No. 1.4e-10;
Matches 81; Conservative 55; Mismatches 134; Indels 30; Gaps 7;
QY 3 KIAFLPAGGAGQYVNGRDFYQYPIVKEITDRASQVVG-----YDLRYLIDTEE 52
DB 883 KIVACGGGQYVNGRDFYQYPIVKEITDRASQVVG-----YDLRYLIDTEE 52
QY 53 DKLNQRYTPALATSVATVAILLOEKGYQPDVAGLSIGYVSALVAGSALDDEDAVALY 112
DB 943 DLL--PVIVQLAAVLSLEVMALGNLGSFGKPSAVIGHSLGEYALYISGVLSAADTYLV 1000
QY 113 AKRGAYMEEAPADSGKMWAVLNTPEVIEEACQKASELGVTVPANTYNTPAQIVAGEVV 172
DB 1001 GMRKLLQRCQGRGTHAMLAVRASPVTLCEVLAESENCEV-----ACHNGPNDTVLSGFLK 1055
QY 173 AYDRAVELLOEAGAKRLPLKVSQGFPHALLLEPASQKLAETLAQVSFSDFTCPVUNGTEA 232
DB 1056 EYVNLQNSASGIGKTL-LKLPFAFHAQVQPILEEFKNVARGVTTHKQIPVLSPLLV 1114
QY 233 AVMKEDIAQ--LLTRQVKEPVRFYESIGVWQEGAGISNF-----TEIGPGKVLISFVK 283
DB 1115 KVIDEKGTVDVYLAHRCREPKVM--VSVLEHARDQHIITDRIVIVDVGKALMAGNIK 1171
RESULT 9
WA_EMBL ID WA_EMBL STANDARD; PRT; 1986 AA.
AC Q03149;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Conidial green pigment synthase (EC 2.3.1.-).
GN WA.


```
Db 1085 RACSEKGLKSTI-LTVYAFHSAQVEFIELEDLQALQGITFNKPSVPFVSALLGEVITEA 1143
Qy 239 DI-----AQLLTROVKEPVRVESIGVMQEAAGISN-----FIEIGPGKVLGSGFVK 283
Db 1144 GSNILNAEYLVHRCRETNVNLSAFEAVRNAKLGGDQTLMEVGPHTVCSGMWK 1196

RESULT 10
PKSI ASPPA
ID PKSI ASPPA STANDARD; PRT; 2109 AA.
AC Q12053;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Aflatoxin biosynthesis polyketide synthase (PKS).
GN PKSL1.
OS Aspergillus parasiticus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5067;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2999;
RX MEDLINE=96042102; PubMed=7592391;
RA Feng G.H., Leonard T.J.;
RT "Characterization of the polyketide synthase gene (pkSL1) required for aflatoxin biosynthesis in Aspergillus parasiticus."
RL J. Bacteriol. 177:6246-6254(1995).
CC -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF AFLATOXIN FROM HEXANOYL COA AND SEVEN MALONATES.
CC -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETHEINE (BY SIMILARITY).
CC -!- PATHWAY: Aflatoxin biosynthesis; first step.
CC -!- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L42766; AAC41675.1; -.
DR EMBL; L42765; AAC41674.1; -.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Peantne attach.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00109; ketoacyl-synt; 1.
DR Pfam; PF00550; pp-binding; 1.
DR Pfam; PF00698; Acyl trans; 1.
DR Pfam; PF00975; Thioesterase; 1.
DR Pfam; PF02801; ketoacyl-synt C; 1.
DR PROSITE; PS00075; ACP_DOMAIN; 1.
KW Transferase; Acyltransferase; Phosphopantetheine; Multifunctional enzyme.
FT DOMAIN 374 805 BETA-KETOACYL SYNTHASE.
FT DOMAIN ? ? ACYL/MALONYL TRANSFERASES.
FT DOMAIN 1714 1785 ACYL CARRIER (ACP).
FT DOMAIN ? 2109 THIOESTERASE.
FT ACT_SITE 543 543 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT ACT_SITE 993 993 ACYL/MALONYL TRANSFERASES (BY SIMILARITY).
FT BINDING 1746 1746 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT ACT_SITE 1937 1937 THIOESTERASE (BY SIMILARITY).
SQ SEQUENCE 2109 AA; 230715 MW; CB701372A16D8551 CRC64;

Query Match 16.7%; Score 254; DB 1; Length 2109;
Best Local Similarity 28.9%; Pred. No. 1.4e-09;
Matches 87; Conservative 48; Mismatches 132; Indels 34; Gaps 11;
-Qy 3 KTAFLAQCGAQLGMGRDFYDQYPIVKETI---DRASQVLGDLRYLI-----DTEED 53
```

```
Db 900 KILFAFTGGGQYATMGQKVYDAYPSFREDLEKFDRLAQSHGFPFSLHVCTSPKGDVEE- 958
Qy 54 KLNQTRVYTPAILATSAIVAILRLLQKGYQYQDMVAGLSIGBYVSALVAGSALDFFEDAVALVA 113
Db 959 --MAPVVQLAITCLQMALTNLMTSFGURPDVTGVHSLGFEAALYAAGVLSASDVVYLVG 1016
Qy 114 KRGAYMEEAPADSGKMVAVLNTPVEVIEACQASSELGVVTPANYTTPAQIVTAGEVVA 173
Db 1017 QRABELQERCQRGTHAMLVAKATPEALSOWIQDHCEVACI-----NGPEDTVLSGGTKN 1071
Qy 174 VDRAVELLQEAQAKRLIPLKVGSGPHTALLEPASPQKLAETLAQ-VSFSD-----FTCPVLVG 228
Db 1072 VAEVQRAMTDNGIKCTL-LKLPFAFSAQVQPIILDDF-EALAQAQATKAPQLLILSPLLR 1129
Qy 229 NT--EAAVMQKEDIAQLLTROVKEPVRVESIGVMQEAAGISN-----FIEIGPGKVLGSGFV 282
Db 1130 TEIHQGVVTPSYVAQ---HCRHTVDMQAALRSAREKGLIDDKTLVIELGPKPLISGMV 1185
Qy 283 K 283
Db 1186 K 1186

RESULT 11
OL56 STRAT
ID OL56 STRAT STANDARD; PRT; 3519 AA.
AC Q07017;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Oleandomycin polyketide synthase, modules 5 and 6.
GN ORPB.
OS Streptomyces antibioticus.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1990;
RN [1]
RX SEQUENCE FROM N.A.
RA MEDLINE=94150470; PubMed=8107683;
RA Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;
RT "Characterisation of a Streptomyces antibioticus gene encoding a type I polyketide synthase which has an unusual coding sequence."
RL Mol. Gen. Genet. 242:358-362(1994).
CC -!- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN LACTONE RING.
CC -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTHETHEINES.
CC -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L09654; AAA19695.1; -.
DR HSSP; P25715; IMLA.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Peantne attach.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF00550; pp-binding; 2.
DR Pfam; PF00698; Acyl trans; 2.
DR Pfam; PF00975; Thioesterase; 1.
DR Pfam; PF02801; ketoacyl-synt C; 2.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 2.
DR PROSITE; PS00606; B_KETOACYL SYNTHASE; 2.
DR PROSITE; PS00075; ACP_DOMAIN; 2.
KW Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Phosphopantetheine; Multifunctional enzyme; Repeat.
```



```

FT NP_BIND 3142 3157 NADP (KR).
FT BINDING 3448 3448 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT CONFLICT 438 438 R -> A (IN REF. 2).
FT CONFLICT 480 480 T -> S (IN REF. 2).
FT CONFLICT 1241 1241 L -> F (IN REF. 2).
FT CONFLICT 2664 2664 G -> V (IN REF. 2).
SQ SEQUENCE 3567 AA; 374413 MW; EB6284F4738AA0C0 CRC64;

Query Match 16.2%; Score 246; DB 1; Length 3567;
Best Local Similarity 26.5%; Pred. No. 9.6e-09;
Matches 80; Conservative 53; Mismatches 149; Indels 20; Gaps 8;

QY 3 KTAFLPAGAOAYLGMGRDGYDQYPIVKETIDRASQVLGVOLRYLI-----DTEEDKLN 56
DB 2012 KPVLPFGQAGQWGWARDLLESEVFAESMSRCAELSPHDTDKLVDVRGGGDPDPE 2071

QY 57 QTRYTPAILATSVAYRLIQXYQYQDMVAGLSLGEYSALVASGALDFEDAVLVAKRG 116
DB 2072 RVDVLQPVLFISMVSLAELWRAHGVTTPAAVVGHSQGEIAAAHVAGALSLEAAAKVVALRS 2131

QY 117 AYMEERAPADSGKQVAVLNPVEVIEACOKASELGVTTPANNVTQAQIVIAGEVVAVDR 176
DB 2132 QVREL--DQGGKMSV-GASRDELETVLARWD--GRVAVAVNGPSTVWAGPTAELD- 2185

QY 177 AVELLOEAGAKRLIPLKVSQGF--HTALLEPAPQKLAETIAQVSFSDFTCPVLGNTEAAV 234
DB 2186 --EPPAEAREMKPRIAVRYASHSPVARIEDRLAELGTITAVRGSVPLHSHVTGTVGEV 2243

QY 235 MQKEDI-AQLLTQVKEPVRFYSIGVMQAGISNFIETGPGKVLGFKKIDQTAHLAH 293
DB 2244 IDTSAMDASWYVNRNLRPVLFEQAVRGLVEQGFDTFVESPFPVL---LMAVEETAHAG 2300

QY 294 VE 295
DB 2301 AE 2302

RESULT 13
ERY3_SACER STANDARD; PRT; 3172 AA.
AC Q0133; Q54097; Q99270;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 5 and 6 (EC 2.3.1.94) (ORF 3) (6-
DE deoxyerythronolide B synthase III) (DEBS 3).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Pseudonocardiaceae; Pseudonocardiaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;
RX MEDLINE=91043075; PubMed=2234082;
RA Cortes J., Haydock S.F., Roberts G.A., Bevitt D.J., Leadlay P.F.;
RT "An unusually large multifunctional polypeptide in the erythromycin-
RL producing polyketide synthase of Saccharopolyspora erythraea."
RN Nature 348:176-178(1990).
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
RL biosynthesis."
RN Science 252:675-679(1991).
RP SEQUENCE FROM N.A.
RX STRAIN=NRRL 2338;
RA MEDLINE=92155230; PubMed=1740151;
RA Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.

```

```

RT Cloning of the structural gene, sequence analysis and inferred domain
RT structure of the multifunctional enzyme.";
RL Eur. J. Biochem. 204:39-49(1992).
CC -|- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC deoxyerythronolide B.
CC -|- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -|- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC BIOSYNTHESIS.
CC -|- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC RESPECTIVELY.
CC -|- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC OF THE FULL-LENGTH CHAIN.
CC -|- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC -|- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X56107; CAA39583.1; -.
CC EMBL; M63677; AAA26495.1; -.
CC EMBL; X62569; CAA44449.1; -.
CC HSPF; P00101; 1CCH.
CC InterPro; IPR002198; ADH_short.
CC InterPro; IPR001227; Ac transferase.
CC InterPro; IPR000794; Ketoacyl-synt.
CC InterPro; IPR003880; Ppantne attach.
CC InterPro; IPR001031; Thioesterase.
CC Pfam; PF00105; adh_short; 1.
CC Pfam; PF00109; ketoacyl-synt; 2.
CC Pfam; PF00550; pp-binding; 2.
CC Pfam; PF00698; Acyl trans; 2.
CC Pfam; PF00975; Thioesterase; 1.
CC Pfam; PF02801; ketoacyl-synt C; 2.
CC PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
CC PROSITE; PS00606; B_KETOACYL SYNTHASE; 2.
CC PROSITE; PS50075; ACP_DOMAIN; 2.
KW Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW Phosphopantetheine; Multifunctional enzyme.
FT DOMAIN 1 1484 MODULE 5.
FT DOMAIN 1485 3172 MODULE 6.
FT DOMAIN 37 484 BETA-KETOACYL SYNTHASE 1.
FT DOMAIN 554 878 ACYLTRANSFERASE (AT) 1.
FT DOMAIN 1116 1298 BETA-KETOACYL REDUCTASE 1.
FT DOMAIN 1394 1464 ACYL CARRIER (ACP) 1.
FT DOMAIN 1488 1954 BETA-KETOACYL SYNTHASE 2.
FT DOMAIN 2021 2335 ACYLTRANSFERASE (AT) 2.
FT DOMAIN 2555 2735 BETA-KETOACYL REDUCTASE 2.
FT DOMAIN 2821 2891 ACYL CARRIER (ACP) 2.
FT DOMAIN 2926 3172 THIOESTERASE.
FT ACT_SITE 199 199 THIOESTER BOND.
FT ACT_SITE 643 643 ACYL-ENZYME INTERMEDIATE.
FT NP_BIND 1118 1164 NADP.
FT BINDING 1427 1427 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT ACT_SITE 1661 1661 THIOESTER BOND.
FT ACT_SITE 2112 2112 ACYL-ENZYME INTERMEDIATE.
FT NP_BIND 2557 2605 NADP.
FT BINDING 2854 2854 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT CONFLICT 231 231 S -> SA (IN REF. 2).
FT CONFLICT 240 240 MISSING (IN REF. 2).

```

```
FT CONFLICT 289 A -> R (IN REF. 2).
FT CONFLICT 493 P -> R (IN REF. 2).
FT CONFLICT 517 PEPRNSLRDTGFTLATRASAMEHRA -> ASRGTRCATPVS
FT 510 RWPDAADWEO (IN REF. 1).
FT 513 A -> R (IN REF. 2).
FT 525 M -> W (IN REF. 2).
FT 536 E -> D (IN REF. 2).
FT 551 R -> G (IN REF. 2).
FT 553 GNPS -> ARIR (IN REF. 2).
FT 673 R -> A (IN REF. 2).
FT 716 MISSING (IN REF. 2).
FT 734 AHK -> GIT (IN REF. 2).
FT 896 R -> RQR (IN REF. 2).
FT 896 R -> RELPVFORQR (IN REF. 1).
FT 994 GVAAYPH -> VLSLRD (IN REF. 2).
FT 1108 RTHPLEA -> ARTRSPR (IN REF. 2).
FT 1126 MISSING (IN REF. 1).
FT 1132 L -> V (IN REF. 2).
FT 1132 A -> R (IN REF. 2).
FT 1134 MISSING (IN REF. 2).
FT 1277 AA -> RR (IN REF. 2).
FT 1278 LQDRE -> STAER (IN REF. 2).
FT 1390 MISSING (IN REF. 2).
FT 1485 G -> R (IN REF. 2).
FT 1518 V -> L (IN REF. 2).
FT 1601 LP -> FA (IN REF. 2).
FT 1725 Q -> L (IN REF. 2).
FT 1732 GBAEG -> ARRA (IN REF. 2).
FT 1739 T -> S (IN REF. 2).
FT 1762 D -> DGAD (IN REF. 2).
FT 2252 QSP -> AVA (IN REF. 2).
FT 2277 G -> GR (IN REF. 2).
FT 2408 LA -> S (IN REF. 2).
FT 2421 NA -> TH (IN REF. 2).
FT 2444 A -> G (IN REF. 2).
FT 2596 P -> A (IN REF. 2).
FT 2609 RRAEGRAA -> AVAKVRR (IN REF. 1).
FT 2721 D -> E (IN REF. 2).
FT 2754 D -> E (IN REF. 2).
SQ SEQUENCE 3172 AA; 331474 MW; DBBD5094E77DD5F CRC64;

Query Match 16.0%; Score 242.5; DB 1; Length 3172;
Best Local Similarity 29.2%; Pred. No. 1.4e-08;
Matches 84; Conservative 47; Mismatches 134; Indels 23; Gaps 10;

QY 3 KTAFLPAGCAQYLGNGRDFYQYPIVKETIDRASQVLG---YDLRYLIDTEEDKLNQ 58
DB 553 RVAMVFGGAGWQMGWARDLLRESQVFADSDTCERALAPHVDSLTLLSGAR-FLDRV 611
QY 59 RYTPAILATSVAIYRLQEKYQPDPMVAGSLGEYSALVASGALDFEDAVLAKRGAY 118
DB 612 DVVQPALFAMVMSLAALWRSHGVEPAAVVGHSGEIAAAHVAGALTLEDAKLVAVRSRV 671
QY 119 MEEAPADSGKMAVLNTPVEVIERACOKASEL-GVVTPTNYNP-AQIVIAGEVAVDR 176
DB 672 LARLG-GQGMAFGLGT-----EQAEIRGFAGALSIVNGPSVVVAGSGPLDE 725
QY 177 AV-ELLQEAGAKRLIPLKVPSPFHTALLEPASQKLAETLAQVSFDFCTPLVGNTEAAVM 235
DB 726 LIAECEAEAKKARIPVDVAS--HSPQVESLEELLTELAGISPVSDVALVSTTG---780
QY 236 QKEDTAQLLT---RQVKEPVRFYSIGVMQAGISNIEIGPGKVL 279
DB 781 QPIDATMDTAYWYANLREOVRFQDQATROLAEAGDAFVVEVSPHPVLT 828

RESULT 14
PPSB MYCTU
ID _PPSB_MYCTU STANDARD; PRT; 1538 AA.
AC Q10978; OS3234;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
```

```
DE Phenolphthiocerol synthesis polyketide synthase ppsB.
GN PF5B OR RV2932 OR MT3002 OR MTCY338.21 OR MTV011.01.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RC MEDLINE=98295987; PubMed=9634230;
RX Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaa F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornby T., Jørgens K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.P., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Kachner A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC SYNTHESIS.
CC -1- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETINE (BY
CC similarity).
CC -1- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DB EMBL; AL021070; GAA15929.1; --
DB EMBL; AE007122; AAK47329.1; --
DB TIGR; MT3002; --
DB TubercuList; RV2932; --
DB InterPro; IPR001227; Ac transferase.
DB InterPro; IPR00794; Ketoacyl-synt.
DB InterPro; IPR003880; Ppantne attach.
DB Pfam; PF00109; ketoacyl-synt; 1.
DB Pfam; PF00550; PP-binding; 1.
DB Pfam; PF00698; Acyl-transf; 1.
DB Pfam; PF02801; ketoacyl-synt-C; 1.
DB PROSITE; PS00606; B KETOACYL SYNTHASE; 1.
DB PROSITE; PS00012; PHOSPHOPANTHETINE; FALSE_NEG.
DB PROSITE; PS00075; ACP DOMAIN; 1.
KW Hypothetical protein; Multifunctional enzyme; Oxidoreductase;
KW Transferrase; NADP; Phosphopantetheine; Complete proteome.
FT DOMAIN 1425 1495 ACYL CARRIER (ACP)
FT BINDING 1458 1458 PHOSPHOPANTHETINE (POTENTIAL).
SQ SEQUENCE 1538 AA; 162527 MW; B55E2A2042AD00CC CRC64;

Query Match 14.8%; Score 224; DB 1; Length 1538;
Best Local Similarity 27.0%; Pred. No. 1.1e-07;
Matches 83; Conservative 55; Mismatches 129; Indels 40; Gaps 13;

QY 4 TAFLPAGCAQYLGNGRDFYQYPIVKETIDRASQVL-----GYDRLYLIDTEEDKLNQ 59
DB 560 TVFYISGRGQWAGMGRCQLLADEFAAAVALEPFFVEQAGFSRLDVATGKE-LVGIE 618
```

QY 60 YTOPAILATSVAIYRLLEQKGYQDMVAGLSGLGYSALVAGSALDGEDAVALVAKGAYM 119
 Db 619 QIQLGLIGMQLTUTELWRSYGVQPDVLVIGHSMGEVAAVAGALTPEAGLRVTATRARLM 678
 QY 120 EEARPADSGWAVANTPVEVIEACOKA--SELGVVTPANTYNTPAQIIVAGEVAVDRA 177
 Db 679 ---APLSGGQGMALLG-----LDAATEALIAIDYPOVTGVIYNSPRQTVIAGTQID-- 728
 QY 178 VELLOEAGAKRLIPLKVS---GPFHTAL--LEPASCKLAETLAQVSFSDF--CPLVG-- 228
 Db 729 -ELIARVRAQNRFPASRVNIEVAPHNPAMDALQPAERS-----ELADLTPTPTIGII 779
 QY 229 NTEAANVOKEDI--AQLLTQVKEPVRFYESI---GVMOEAGISNFIETIGPKVLSGFVK 283
 Db 780 STTYADLHTQIPDAEHWATNMNRPVRFQAIASAGSGADGAYHTFIETISAPHLTQAIA 839
 QY 284 KIDQTAH 290
 Db 840 DTLEDAH 846

RESULT 15
 PPSSA MYCTU STANDARD; PRT; 1876 AA.
 AC Q10977;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phenolphthiocerol synthesis polyketide synthase ppsa.
 GN PPSA OR RV2931 OR MT3000 OR MTCY338.20.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RA MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekalia F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
 CC SYNTHESIS.
 CC -1- COPACITOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES (By
 CC similarity).
 CC -1- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z74697; CAA98988.1; -
 DR EMBL; AE007122; AAK47328.1; -
 DR TIGR; MT3000; -
 DR Tuberculist; RV2931; -
 DR Interpro; IPR001227; Ac_transferase.
 DR Interpro; IPR00794; Ketoacyl-synt.
 DR Interpro; IPR003880; Pantne_attach.
 DR Pfam; PF00109; ketoacyl-synt; 1.
 DR Pfam; PF00550; pp-binding; 2.
 DR Pfam; PF00698; Acyl trans; 1.
 DR Pfam; PF02801; ketoacyl-synt C; 1.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
 DR PROSITE; PS00608; E_KETOACYL_SYNTHASE; 1.
 DR PROSITE; PS00075; ACP_DOMAIN; 2.
 KW Multifunctional enzyme; Oxidoreductase; Transferase; NADP; Repeat;
 KW Phosphopantetheine; Complete proteome.
 FT DOMAIN 7 80 ACYL CARRIER (ACP) 1.
 FT DOMAIN 1764 1833 ACYL CARRIER (ACP) 2.
 FT NP_BIND 1498 1503 NADP (POTENTIAL).
 FT BINDING 62 62 PHOSPHOPANTETHEINE (POTENTIAL).
 FT ACT_SITE 273 273 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
 FT ACT_SITE 720 720 MALONYLTRANSFERASE (BY SIMILARITY).
 FT BINDING 1796 1796 PHOSPHOPANTETHEINE (POTENTIAL).
 FT CONFLICT 624 624 D -> E (IN REF. 2).
 FT CONFLICT 877 877 R -> H (IN REF. 2).
 FT CONFLICT 1323 1323 G -> S (IN REF. 2).
 SQ SEQUENCE 1876 AA; 198834 MW; D9783BDB48792110 CRC64;

Query Match 13.7%; Score 208.5; DB 1; Length 1876;
 Best Local Similarity 26.4%; Pred. No. 1.6e-06;

Matches 81; Conservative 52; Mismatches 135; Indels 39; Gaps 13;

QY 4 TAPLFAQGGAYLGMGRDFYDQYPIVKTIDRASQVL----GYDLRYLIDTEEDKLNQTR 59
 Db 631 TVFVYSGRGSQWAGNQGLLADEPAFAAAVAELEPVFEVGEQAGSLHVDVLANGE-LVGIE 689

QY 60 YTOPAILATSVAIYRLLEQKGYQDMVAGLSGLGYSALVAGSALDGEDAVALVAKGAYM 119
 Db 690 QIQLGLIGMQLTUTELWRSYGVQPDVLVIGHSMGEVAAVAGALTPEAGLRVTATRARLM 749

QY 120 EEARPADSGWAVANTPVEVIEACOKASELGVVTPANTYNTPAQIIVAGEVAVDRAVE 179
 Db 750 ---APLSGGQGMALLDAPTEAL---IADFPQVTIGIYNSPRQTVIAGTQID---E 800

QY 180 LLOEAGAKRLIPLKVS---GPFHTAL--LEPASCKLAETLAQVSFSDF--CPLVG--NT 230
 Db 801 LIARVRAQNRFPASRVNIEVAPHNPAMDALQPAERS-----ELADLTPTPTIGIIST 852

QY 231 EAAVWOKEDI--AQLLTQVKEPVRFYESI---GVMOEAGISNFIETIGPKVLSGFVKI 285
 Db 853 TYADLHTQIPVDAEHWATNMNRPVRFQAIASAGSGADGAYHTFIETISAPHLT---QAI 909

QY 286 DQTAHLA 292
 Db 910 IDTLHSA 916

Search completed: June 11, 2003, 20:03:40
 Job time : 39 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 19:58:52 ; Search time 59 Seconds
(without alignments)
1068.651 Million cell updates/sec

Title: US-09-308-397-2
Perfect score: 1518
Sequence: 1 MKTAFUFGAGQAYLNGR.....QTAHLAVEDQASIVALLEK 306

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL 21.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phase.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1518	100.0	306	Q9FBC4	Q9fbc4 streptococ
2	1048	69.0	308	16 Q9CHP8	Q9chf8 lactococcus
3	1036	68.2	312	16 Q99YD5	Q99y45 streptococ
4	775.5	51.1	309	16 Q9R9V9	Q9r9v9 thermoaer
5	713.5	47.0	308	16 Q97DA5	Q97da5 clostridium
6	703.5	46.3	314	16 Q9XLM2	Q9xlm2 clostridium
7	657.5	43.3	312	16 Q9KQK6	Q9kqk6 vibrio chol
8	654	43.1	313	16 Q9KA02	Q9ka02 bacillus ha
9	640.5	42.2	309	16 Q9X817	Q9x817 escherichia
10	628.5	41.4	309	16 Q9ZFF6	Q9zff6 yersinia pe
11	626	41.2	312	16 Q9CJ57	Q9cjs7 pasteurella
12	612.5	40.3	309	16 Q9ZL75	Q9zlj5 salmonella
13	611.5	40.3	307	2 Q9ZA34	Q9za34 vibrio mari
14	598	39.4	309	16 Q9Y0J3	Q9y0j3 ralbstonia s
15	594.5	39.2	326	16 Q67041	Q67041 aquifex aeo
16	588.5	38.8	313	16 Q9Y689	Q9y689 listeria mo

17	586	38.6	313	16 Q92AK0	Q92ak0 listeria in
18	570	37.5	299	16 Q8RGX6	Q8rgx6 fusobacteri
19	570	37.5	308	16 Q9JWS8	Q9jws8 naissieriam
20	566	37.3	308	16 Q9JXR4	Q9jxr4 naissieriam
21	556.5	36.7	312	16 Q54437	Q54437 pseudomonas
22	551.5	36.3	308	16 Q99UN8	Q99un8 staphylococ
23	550.5	36.3	308	16 Q93QD4	Q93qd4 staphylococ
24	533.5	35.1	317	16 Q9PF17	Q9pf17 xylella fas
25	531	35.0	308	16 Q9JSG5	Q9jsg5 chlamydia p
26	529	34.8	305	16 Q9RT24	Q9rt24 deinnococcus
27	528	34.8	308	16 Q9Z8P1	Q9z8p1 chlamydia p
28	526	34.7	293	16 Q9WZ05	Q9wzq5 thermotoga
29	503.5	33.2	400	2 Q45564	Q45564 bacillus su
30	492	32.4	331	16 Q8UGE4	Q8uge4 agrobacteri
31	485	31.9	314	16 Q9YFP2	Q9yfp2 bruceella me
32	482	31.8	314	16 Q9KGI1	Q9kgt1 rhizobium m
33	479.5	31.6	400	2 Q9J157	Q9j157 bacillus su
34	479	31.6	292	16 Q8Z061	Q8z061 anabaena sp
35	479	31.6	314	16 Q9A7P6	Q9a7p6 caulobacter
36	471.5	31.1	1110	2 Q8RL73	Q8rl73 pseudomonas
37	468	30.8	314	16 Q92GK7	Q92gk7 rickettsia
38	464.5	30.6	308	16 Q9PKF6	Q9pkf6 chlamydia m
39	461.5	30.4	306	16 Q9PJ11	Q9pj11 campylobact
40	461	30.4	315	16 Q9E4T5	Q9e4t5 rhizobium l
41	457.5	30.1	400	2 Q9R9J2	Q9r9j2 bacillus su
42	451.5	29.7	308	16 Q84241	Q84241 chlamydia t
43	445	29.3	319	16 Q9ZCJ6	Q9zcj6 rickettsia
44	440.5	29.0	349	10 Q9XGMO	Q9xgm0 brassica na
45	439.5	29.0	293	10 Q8RU07	Q8ru07 arabidopsis

ALIGNMENTS

RESULT 1

Q9FBC4	PRELIMINARY;	PRT;	306 AA.
ID	Q9FBC4		
AC	Q9FBC4;		
DT	01-MAR-2001 (TREMREL. 16, Created)		
DT	01-MAR-2001 (TREMREL. 16, Last sequence update)		
DT	01-JUN-2002 (TREMREL. 21, Last annotation update)		
DE	Malonyl-CoA:ACP transacylase (Malonyl CoA-acyl carrier protein transacylase).		
DE	FABD OR SP0420.		
GN	Streptococcus pneumoniae.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;		
OC	Streptococcaceae; Streptococcus.		
OX	NCBI_TaxID=1313;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=R6;		
RX	MEDLINE=20365714; PubMed=10910344;		
RT	Heath R.J., Rock C.O.;		
RT	"A tetracycline-resistant bacterial enzyme.";		
RL	Nature 406:145-146(2000).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=11GR4;		
RX	MEDLINE=21357209; PubMed=11463916;		
RA	Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,		
RA	Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,		
RA	Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,		
RA	Umayah L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,		
RA	Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,		
RA	McDonald L.E., Feldblyum T.V., Angiolis S., Dickinson T.C., Hickey E.K.,		
RA	Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,		
RA	Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;		
RT	"Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.";		
RL	Science 293:498-506(2001).		
DR	EMBL; AF197933; AAF98274.1;		
DR	EMBL; A6007354; AAK74563.1;		
DR	HSSP; P25715; INLA.		

```

DR TIGR; SPO420; -
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR004410; FabD.
DR Pfam; PF00698; Acyl_transf; 1.
DR TIGRfams; TIGR00128; fabd; 1.
KW Complete proteome.
SQ SEQUENCE 306 AA; 33091 MW; 83324189A89C2464 CRC64;

Query Match 100.0%; Score 1518; DB 16; Length 306;
Best Local Similarity 100.0%; Pred. No. 1.9e-99;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKTAFLAGQAQGLGMGRDFDYQYPIVKETIDRASQVLGYDLRLYLIDTEEDKLNQTRY 60
DB 1 MTKTAFLAGQAQGLGMGRDFDYQYPIVKETIDRASQVLGYDLRLYLIDTEEDKLNQTRY 60
QY 61 TQPAILATSVAIYRLLOEKGYQPDWVAGLSGYSALVAGSALDFFEDAVALKRQAYME 120
DB 61 TQPAILATSVAIYRLLOEKGYQPDWVAGLSGYSALVAGSALDFFEDAVALKRQAYME 120
QY 121 EAPADSGKQWAVLNTPEVIEEACQKASEL--GVVTPANTYNTPAQIVIAGEVAVDRAVEL 180
DB 121 EAPADSGKQWAVLNTPEVIEEACQKASELGVVTPANTYNTPAQIVIAGEVAVDRAVEL 180
QY 179 ELLOEAGAKRLIPLKVGSPFHTALLESASOKLAETLAQVSFSDFTCPVLGNTAEAVMOKEDI 240
DB 179 ELLOEAGAKRLIPLKVGSPFHTALLESASOKLAETLAQVSFSDFTCPVLGNTAEAVMOKEDI 240
QY 239 DIAQLTRQVKPEPVRYEYSGVQWQAGISNFIETGPGKVLGSGFVKKIDQTAHLAHVEDQA 298
DB 239 DIAQLTRQVKPEPVRYEYSGVQWQAGISNFIETGPGKVLGSGFVKKIDQTAHLAHVEDQA 298
QY 241 EVKGLLRQWKEPVRYEYSGVQWQAGISNFIETGPGKVLGSGFVKKIDQTAHLAHVEDQA 300
DB 241 EVKGLLRQWKEPVRYEYSGVQWQAGISNFIETGPGKVLGSGFVKKIDQTAHLAHVEDQA 300
QY 301 VALLEK 306
DB 301 VALLEK 306

RESULT 2
Q9CHF8 PRELIMINARY; PRT; 308 AA.
AC Q9CHF8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative malonyl CoA-acyl carrier protein transacylase
DE (EC 2.3.1.39).
GN FABD OR L1403.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauer S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RA "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
DR EMBL; AE006311; AAK04871.1; -.
DR HSP; P25715; 1MLA.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR004410; FabD.
DR Pfam; PF00698; Acyl_transf; 1.
DR TIGRfams; TIGR00128; fabd; 1.
DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN 1.
KW Complete proteome.
SQ SEQUENCE 308 AA; 33541 MW; BAB129A7A9EE4CF CRC64;

Query Match 69.0%; Score 1048; DB 16; Length 308;
Best Local Similarity 67.5%; Pred. No. 3e-66;
Matches 308; Conservative 40; Mismatches 58; Indels 2; Gaps 1;

QY 1 MTKTAFLAGQAQGLGMGRDFDYQYPIVKETIDRASQVLGYDLRLYLIDTEEDKLNQTRY 60
DB 1 MTKTAFLAGQAQGLGMGRDFDYQYPIVKETIDRASQVLGYDLRLYLIDTEEDKLNQTRY 60
QY 61 TQPAILATSVAIYRLLOEKGYQPDWVAGLSGYSALVAGSALDFFEDAVALKRQAYME 120
DB 61 TQPAILATSVAIYRLLOEKGYQPDWVAGLSGYSALVAGSALDFFEDAVALKRQAYME 120
QY 121 EAPADSGKQWAVLNTPEVIEEACQKASELGVVTPANTYNTPAQIVIAGEVAVDRAVEL 180
DB 121 EAPADSGKQWAVLNTPEVIEEACQKASELGVVTPANTYNTPAQIVIAGEVAVDRAVEL 180
QY 179 ELLOEAGAKRLIPLKVGSPFHTALLESASOKLAETLAQVSFSDFTCPVLGNTAEAVMOKEDI 240
DB 179 ELLOEAGAKRLIPLKVGSPFHTALLESASOKLAETLAQVSFSDFTCPVLGNTAEAVMOKEDI 240
QY 239 DIAQLTRQVKPEPVRYEYSGVQWQAGISNFIETGPGKVLGSGFVKKIDQTAHLAHVEDQA 298
DB 239 DIAQLTRQVKPEPVRYEYSGVQWQAGISNFIETGPGKVLGSGFVKKIDQTAHLAHVEDQA 298
QY 241 EVKGLLRQWKEPVRYEYSGVQWQAGISNFIETGPGKVLGSGFVKKIDQTAHLAHVEDQA 300
DB 241 EVKGLLRQWKEPVRYEYSGVQWQAGISNFIETGPGKVLGSGFVKKIDQTAHLAHVEDQA 300
QY 301 VALLEK 306
DB 301 VALLEK 306

RESULT 3
Q99YD5 PRELIMINARY; PRT; 312 AA.
AC Q99YD5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative malonyl CoA-acyl carrier protein transacylase
DE (EC 2.3.1.39).
GN FABD OR SPV1750.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Perrett J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RA "Complete genome sequence of an M1 strain of Streptococcus pyogenes."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006603; AAK34494.1; -.
DR HSP; P25715; 1MLA.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR000315; BPD_transp.
DR InterPro; IPR004410; FabD.
DR Pfam; PF00698; Acyl_transf; 1.
DR TIGRfams; TIGR00128; fabd; 1.
DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN 1.
KW Complete proteome.
SQ SEQUENCE 312 AA; 34214 MW; OBAC6A9D9199D5B CRC64;

Query Match 68.2%; Score 1036; DB 16; Length 312;
Best Local Similarity 68.3%; Pred. No. 2.2e-65;
Matches 209; Conservative 39; Mismatches 58; Indels 0; Gaps 0;

QY 1 MTKTAFLAGQAQGLGMGRDFDYQYPIVKETIDRASQVLGYDLRLYLIDTEEDKLNQTRY 60
DB 1 MTKTAFLAGQAQGLGMGRDFDYQYPIVKETIDRASQVLGYDLRLYLIDTEEDKLNQTRY 60
QY 61 TQPAILATSVAIYRLLOEKGYQPDWVAGLSGYSALVAGSALDFFEDAVALKRQAYME 120
DB 61 TQPAILATSVAIYRLLOEKGYQPDWVAGLSGYSALVAGSALDFFEDAVALKRQAYME 120
QY 121 EAPADSGKQWAVLNTPEVIEEACQKASELGVVTPANTYNTPAQIVIAGEVAVDRAVEL 180
DB 121 EAPADSGKQWAVLNTPEVIEEACQKASELGVVTPANTYNTPAQIVIAGEVAVDRAVEL 180
QY 179 ELLOEAGAKRLIPLKVGSPFHTALLESASOKLAETLAQVSFSDFTCPVLGNTAEAVMOKEDI 240
DB 179 ELLOEAGAKRLIPLKVGSPFHTALLESASOKLAETLAQVSFSDFTCPVLGNTAEAVMOKEDI 240
QY 239 DIAQLTRQVKPEPVRYEYSGVQWQAGISNFIETGPGKVLGSGFVKKIDQTAHLAHVEDQA 298
DB 239 DIAQLTRQVKPEPVRYEYSGVQWQAGISNFIETGPGKVLGSGFVKKIDQTAHLAHVEDQA 298
QY 241 EVKGLLRQWKEPVRYEYSGVQWQAGISNFIETGPGKVLGSGFVKKIDQTAHLAHVEDQA 300
DB 241 EVKGLLRQWKEPVRYEYSGVQWQAGISNFIETGPGKVLGSGFVKKIDQTAHLAHVEDQA 300
QY 301 VALLEK 306
DB 301 VALLEK 306

```

```
Db 121 EAAPOGSGKVAAMNTDVQVIEEVQIAAKHGVAAPANVTNTPSIVIGQTDVAVVEL 180
Qy 181 LQAGAKRLIPLKVSQPFHTALLEPASQKLAETLAQVSPDFTCPVGNTEAAVMQKEDI 240
Db 181 LKRGVYKRIPLNVSGPFHTALLEPASRLLAELERYNFSDFKPLVNGNTEANIMEKDI 240
Qy 241 AQLTRQVKEPVPFYESIGVMOEAGISNFIETGPGKVLGFGFKKIDQTAHLAHVEDQASL 300
Db 241 PELLARQVNEPVRFDVSATLVESGITQFIEVGPKVLGTGFKKIDKLLCTSVENWVSL 300
Qy 301 VALLEK 306
Db 301 RLFLDR 306

RESULT 4
Q8R9V9 PRELIMINARY; PRT; 309 AA.
AC Q8R9V9;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE (acyl-carrier-protein) S-malonyltransferase.
GN FABD OR TTE1473.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4T / JCM11007;
RX MEDLINE=21997336; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AB013105; AAM24695.1; -.
KW Transferase; Complete proteome.
SQ SEQUENCE 309 AA; 33891 MW; 81557C60A4AC7983 CRC64;

Query Match 51.1%; Score 775.5; DB 16; Length 309;
Best Local Similarity 50.8%; Pred. No. 5.4e-47;
Matches 154; Conservative 58; Mismatches 90; Indels 1; Gaps 1;

Qy 3 KTAFLPAGGAGYGLMGGRDFYDQYPIVKETIDRASQVLGYDLRYL-IDTEEDKLNQTRYT 61
Db 2 KTAFLPAGGAGYAGMGKGIYEKYEAKEIFERADEALGFNISKLCFSGPBEELMKENT 61
Qy 62 QPAILATSVAIYRLIQEKYQPDWAGLSLGEYSALVSGALDPEDAVLVAKRGAYMEE 121
Db 62 QPAILTVSVALTRVLQKRGKPDVTAGLSLGEYSLSLVAELDFEDAVLVKRGKYMQE 121
Qy 122 AAPADSGKVAAMNTDVQVIEEVQIAAKHGVAAPANVTNTPSIVIGQTDVAVVEL 181
Db 122 VVPEGVGTMAAILGLPNEVEIEICIASVGVVEPANYNCPQLVVSQVAVRAVELA 181
Qy 182 QPAGAKRLIPLKVSQPFHTALLEPASQKLAETLAQVSPDFTCPVGNTEAAVMQKEDIA 241
Db 182 KERGAKAVVLAUSAFFHCSMLKAGELLAKELDKLVKVPVINSVTADYVQKQVK 241
Qy 242 QLLTRQVKEPVPFYESIGVMOEAGISNFIETGPGKVLGFGFKKIDQTAHLAHVEDQASL 301
Db 242 ELLIKQVSHVLEWQSVRNIEDGVDVTFIEIGFGKTLGSGFVKKIDRSTVNLNFEDESLM 301
Qy 302 ALL 304
Db 302 KAL 304

RESULT 5
Q97DAS PRELIMINARY; PRT; 308 AA.
ID Q97DAS
```

```
Q97DAS;
AC 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Malonyl CoA-acyl carrier protein transacylase.
GN CAC3575.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Ghu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007854; AAK81498.1; -.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR004410; FabD.
DR Pfam; PF00698; Acyl transf. 1.
DR TIGRPFAMs; TIGR00128; fabD; 1.
KW Complete proteome.
SQ SEQUENCE 308 AA; 33811 MW; 26901C5C584AD4E2 CRC64;

Query Match 47.0%; Score 713.5; DB 16; Length 308;
Best Local Similarity 48.5%; Pred. No. 1.3e-42;
Matches 147; Conservative 55; Mismatches 97; Indels 3; Gaps 3;

Qy 1 MTKTAPLAGGAGYGLMGGRDFYDQYPIVKETIDRASQVLGYDLRYL-IDTEEDKLNQTRYT 59
Db 1 MGKIAFVFGQGGYGVGMKDLVDNYSQAKETFDKADVLGFKISLCFSGKDEELNLT 60
Qy 60 YTPAILATSVAIYRLIQEKYQPDWAGLSLGEYSALVSGALDPEDAVLVAKRGAY 118
Db 61 NTQPAVLITSIALRALALEEKIKPDVAVGLSLGEYSALVSGSGSFGSFGSFGSFGSFGS 120
Qy 119 MEAAAPADSGKVAAMNTDVQVIEEVQIAAKHGVAAPANVTNTPSIVIGQTDVAVVEL 178
Db 121 MQSAVPKIGTMAAILGLEGVVVRGICAEQSGEIVGVANVYCPQQLVIVAGEVAVESAC 180
Qy 179 ELQEGAKRLIPLKVSQPFHTALLEPASQKLAETLAQVSPDFTCPVGN-TEAAVMQK 237
Db 181 TKLKGSGARTVMSVSGPHTSMLKSAEKLKEELKNINIEDMKVPVITNVTGVDYVEDK 240
Qy 238 EDIAQLTRQVKEPVPFYESIGVMOEAGISNFIETGPGKVLGFGFKKIDQTAHLAHVEDQ 297
Db 241 DOIKGLKKQWSSVRWEDTIRMDGVDVTFIEIGFGKTLSSFIKKIRKMTIFNIEKA 300
Qy 298 ASL 300
Db 301 EDL 303

RESULT 6
Q8XLH2 PRELIMINARY; PRT; 314 AA.
ID Q8XLH2
AC Q8XLH2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Malonyl CoA-acyl carrier protein transacylase.
GN FABD OR CPE1069.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
```

```

RC STRAIN=13 / TYPE A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.,
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RL EMBL; AF003189; BAB0775.1; -.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR004410; FAbD.
DR Pfam; PF00698; Acyl_transf; 1.
DR TIGRFAMs; TIGR00128; fabd; 1.
KW Complete proteome.
SQ SEQUENCE 314 AA; 34273 MW; 801E151B80390156 CRC64;

Query Match 46.3%; Score 703.5; DB 16; Length 314;
Best Local Similarity 47.7%; Pred. No. 6.8e-42;
Matches 147; Conservative 61; Mismatches 97; Indels 3; Gaps 3;

QY 1 MTKTAFLPAGOGAOLVGMGRDYPQYPIVKTIDRASOLGYDLRYLIDTE-EDKLNQTR 59
DB 1 MAKLGFLPAGOGAOLVGMGRDYPQYPIVKTIDRASOLGYDLRYLIDTE-EDKLNQTR 60

QY 60 YTPAILATSVAIYRLLOEKGY-QPDMVAGLSGLGEYSALVAGSALDGFEDAVLVAKRGAYM 119
DB 61 FTQPAIITTNMAILLTALDKLVKSHISGLSLGEYSALIHSGAINPFGVGLVKRGKFM 120

QY 120 EEARAPADSGKMWAVLNPVEVIEACOKASELGWVTPANTYNTPAQIVIAGEVAVDRAVE 179
DB 121 QEVAEGTGGWAVLNPVEVIEACOKASELGWVTPANTYNTPAQIVIAGEVAVDRAVE 180

QY 180 LLOQAGAKRLIPLKVSQPFHTALLEPASQKLAETLAQVSPSDFTCPLVGNTEA-AVMQKE 238
DB 191 FIKEVGG-RAIKLPVSPFFHCSMLQPAEKLDELINSLKNGIWMNVKGEVLEDD 239

QY 239 DIAQLLTQVKEPVPFYESIGVMOEAGISNFIETGPGKVLGSGFVKKIDQTAHLAHVEDQA 298
DB 240 NIELTTSQVKPVLFIENDIEKMTESGVDTEIETGPGKALSGFVKKINKVNTVLANVEDLK 299

QY 299 SLVALLEK 306
DB 300 SLEKTLK 307

RESULT 7
Q9KQH6 PRELIMINARY; PRT; 312 AA.
AC Q9KQH6;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Malonyl CoA-acyl carrier protein transacylase.
GN VC2022.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae";
RL Nature 406:477-483(2000).
DR EMBL; AE004276; AAF95170.1; -.
DR HSSP; P25715; 1MLA.
DR TIGR; VC2022; -.

```

```

DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR004410; FAbD.
DR Pfam; PF00698; Acyl_transf; 1.
DR TIGRFAMs; TIGR00128; fabd; 1.
KW Complete proteome.
SQ SEQUENCE 312 AA; 33217 MW; 458A38C7E1D2356F CRC64;

Query Match 43.3%; Score 657.5; DB 16; Length 312;
Best Local Similarity 47.4%; Pred. No. 1.2e-38;
Matches 146; Conservative 58; Mismatches 97; Indels 7; Gaps 4;

QY 1 MTKTAFLPAGOGAOLVGMGRDYPQYPIVKTIDRASOLGYDLRYLIT-DTEEDKLNQTR 59
DB 6 MSKFAIVFPQGSQAVGMLADLAQYAVVKTFAEASEVLGYDLWALVQDGPVEDLNQTF 65

QY 60 YTPAILATSVAIYRLLOEKGY-QPDMVAGLSGLGEYSALVAGSALDGFEDAVLVAKRGAY 118
DB 66 RTQPALLAASVAIWRVMOQLGLEQPAVLGHSGLGEYSALVAGVDFKQAIKLVELRGQL 125

QY 119 MEERAPADSGKMWAVLNPVEVIEACOKASELGWVTPANTYNTPAQIVIAGEVAVDRAV 178
DB 126 MOQAVPAGTGMAYIIGLEDAIAKACADAAGQGVSPVNFNSPGQVVIAGOKDAVERAG 185

QY 179 ELLQAGAKRLIPLKVSQPFHTALLEPASQKLAETLAQVSPSDFTCPLVGNTEA-AVMQKE 238
DB 186 VLCKEAGAKRALPLPVSVPSCALMKPAADBLAKTLAELEFNAPQIPVNNVD--VVAET 243

QY 239 D---IAQLLTQVKEPVPFYESIGVMOEAGISNFIETGPGKVLGSGFVKKIDQTAHLAHE 295
DB 244 DVKTKDALIKQLVSPVRWTECVQMSAQGVKLEMGFGKVLIGLTKRIIVKTLGVAVN 303

QY 295 DOASLVAL 303
DB 304 DVASLDAV 311

RESULT 8
Q9KA02 PRELIMINARY; PRT; 313 AA.
AC Q9KA02;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39).
GN FABD OR BH2492.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AF001515; BAB06211.1; -.
DR HSSP; P25715; 1MLA.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR000629; DEAD_box.
DR InterPro; IPR004410; FAbD_box.
DR Pfam; PF00698; Acyl_transf; 1.
DR TIGRFAMs; TIGR00128; fabd; 1.
DR PROSITE; PS00039; DEAD_ATP_HELICASE; UNKNOWN_1.
KW Transferase; Acyltransferase; Complete proteome.
SQ SEQUENCE 313 AA; 33585 MW; B009ACA5AD8C3968 CRC64;

Query Match 43.1%; Score 654; DB 16; Length 313;
Best Local Similarity 45.6%; Pred. No. 2.1e-38;
Matches 141; Conservative 58; Mismatches 106; Indels 4; Gaps 4;

```



```

Qy      1  MTKTAPLPAGQAOVLGMRBFDYDQYPIVKSTIDRASQVLGYDL-RVLIIDTEEDKLNQTR 59
Db      1  MAKVAFLPFGQSQSGVMGSELLSEBK-AKEIFDEADERLGYSLSSIMFEGPEKURRTE 59
Qy     60  YTPQALLATSVAIYRLLQEKGYOPDMVAGLSLGEYSALVAGSALDFEDAVALVAKRGAYM 119
Db     60  NTQPALLTMTSTAVLSLVREYGIKPDYTAGHLSGEYSALVAGSLTFADADAYVTHHGLFM 119
Qy    120  EEPAAPADSGMKVAVLNTPEVIEEACQKASELG-VVTPANYNTPAQIVIAAGEVAVVDRAV 178
Db    120  EEAVPFEGEGAMAAILGMRDELEQVTKRVTEAGAVWELANLPCPGQIVISGSAEGVEQAS 179
Qy    179  ELLQEAGAKRLIPLKVSQGFHTALLEPASQKLAETLQOVSEFTECPVLGNTTAAVMQK- 237
Db    180  EBAKAGAKRVIPLQVSGPFHSSLMKPAEKLAVLADLATAAAPETIANVTADIVQKA 239
Qy    238  EDIAQLLTQVKPEPVRVFEISGMQEAGISNFIIEIGPKVLSGFVKKIDOTAHLAHVQD 297
Db    240  ADIRASLIEQVYSPVRWEDTVRMLELGVDFVEIGSVLGLVRKVQRVNVFVSVD 299
Qy    298  ASLVALLEK 306
Db    300  ASIEAMVKK 308

```

```

RESULT 9
Q8X817
ID ID Q8X817 PRELIMINARY; PRT; 309 AA.
AC Q8X817;
DT DT 01-MAR-2002 (trEMBLrel. 20, Created)
DT DT 01-MAR-2002 (trEMBLrel. 20, Last sequence update)
DT DT 01-JUN-2002 (trEMBLrel. 21, Last annotation update)
DE DE Malonyl-CoA-[acyl-carrier-protein] transacylase.
GN FABD OR Z1731 OR ECS1470.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
CX NCBI_TaxID=83334;
CN [1]_
RN SEQUENCE FROM N.A.
RP RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grothbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potancusis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533 (2001).
RN [2]
RP RC SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res 8:11-22(2003).
EL ENBL; AS005319; AAG55638.1; -
DR ENBL; AS005355; BAB34893.1; -
DR InterPro: IPR001227; Ac transferase.
DR InterPro: IPR004410; FabD.
DR Pfam: PF00698; Acyl transf. 1.
DR TIGRfams: TIGR00128; fabD; 1.
KW Complete proteome.
SQ SEQUENCE 309 AA; 32412 MW; 3DCB41120D526F71 CRC64;

```

Matches	141;	Conservative	59;	Mismatches	103;	Indels	7;	Gaps	4;
Qy	1	MTKTAFLEFGGAGYVGMGRDFYDQYPIVKETIDRASQVLYDRLYLIDTEE---	DKLNQ	57					
Db	1	MTQFAFVFPQGGSGTQVGLMDLMAASYPYVEETFAEASAAALGYDLWAL--	TQOQPAEELNK	58					
Qy	58	TRYTQPAILATSVAIYRLLEQKGY--PDMVAGLSLGEYSALVASGALDPEDAVALVAKRG	116						
Db	59	TWOTQPAILTASVALYRVVQVQGGKAPAMVAGHSLGEYSALVCAGVIDFADAVRLVEMRG	118						
Qy	117	AYNEEAPADSGKMWAVLNTPEVIEBEACOKASELGVVTPANNVNTPAQIVIAEYVAVOR	176						
Db	119	KFMQEAAPVPGTGMATAIIGLDDASIGKACEEAAGQVSPVNFSPGQVVIAGHKEAVER	178						
Qy	177	AVELLQEAQAKRLILPKVSGSPRHETALLEPASOKLAETLAQVSPSDFTCPLVGVNTEAAVMQ	236						
Db	179	AGAACAAAGAKRALPLPVSVPSCALMKPAADKLAVELAKITFNATPTPVVNVNDVKCET	238						
Qy	237	KED-IAQLITRQVKEPRVYESIGVMOEAGISNFIPIGPKVLISGVFKKIDQTAHLAHVE	295						
Db	239	NGDAIRDALVRLYNPQVTKSVYVMAAGVEHLXEVGPGKVLITGTIKRIVDTLTASALN	298						
Qy	296	DQASLVALLE 305							
Db	299	EPSMAAALE 308							
RESULT 10									
Q8ZFT6	ID	Q8ZFT6	PRELIMINARY;	PRT;	305	AA.			
AC	Q8ZFT6;								
DT	01-MAR-2002	(TEMBLrel. 20, Created)							
DT	01-MAR-2002	(TEMBLrel. 20, Last sequence update)							
DT	01-JUN-2002	(TEMBLrel. 21, Last annotation update)							
DE	Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39).								
GN	FABD OR YPO1598.								
OS	Yersinia pestis.								
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;								
OC	Yersinia.								
OC	NCBI_TaxID=632;								
RN	[1]								
SEQUENCE FROM N.A.									
RC	STRAIN=CO-92 / BIOVAR ORIENTALIS;								
RX	MEDLINE=21470413; PubMed=11586360;								
RA	Parkhill J., Wren B. W., Thomson N. R., Titball R. W., Holden M. T. G.,								
RA	Prentice M. B., Sebaitia M., James K. D., Churcher C., Mungall K. L.,								
RA	Baker S., Basham D., Bentley S. D., Brooks K., Cerdeno-Farraga A. M.,								
RA	Chillingworth T., Cronin A., Davies R. M., Davis P., Dougan G.,								
RA	Feltwell T., Hamlin N., Holroyd S., Jagels K., Kariyshev A. V.,								
RA	Leather S., Moule S., Oyston P. C. F., Quail M., Rutherford K.,								
RA	Simmonds M., Skelton J., Stevens K., Whitehead S., Bartell B. G.;								
RT	"Genome sequence of Yersinia pestis, the causative agent of plague.";								
RL	Nature 413:523-527(2001).								
DR	EMBL; AJ414149; CAC90420.1; -								
DR	InterPro; IPR001227; Ac transferase.								
DR	InterPro; IPR004410; FabD.								
DR	Pfam; PF00698; Acyl transf. 1.								
DR	TIGRfams; TIGR00128; fabd; 1.								
KW	Transferase; Acyltransferase; Complete proteome.								
SQ	SEQUENCE 309 AA; 32651 MW; 5F090049FF9A848B CRC64;								
Query Match									
Best Local Similarity 41.4%; Score 628.5; DB 16; Length 309;									
Matches 140; Conservative 56; Mismatches 109; Indels 3; Gaps 3;									
Qy	1	MTKTAFLEFGGAGYVGMGRDFYDQYPIVKETIDRASQVLYDRLYLIDT-EEDKLNQTR	59						
Db	1	MSKFMVFPQGGSGSLMDLADLAQAQPIVEATSEASSVLGYDLWQLVQOQPAEELNKTW	60						
Qy	60	YTQPAILATSVAIYRLLEQKGY--PDMVAGLSLGEYSALVASGALDPEDAVALVAKRGAY	118						
Db	61	OTQPAILTASVAIYRWQHGKGLPSLMAGHSLGEYSALVCAGVLDFKQAVSLVEIRGKL	120						

